(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
19 October 2006 (19.10.2006)

PCT

(10) International Publication Number
WO 2006/110264 A2

(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/US2006/009870

(22) International Filing Date: 16 March 2006 (16.03.2006)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/663,014 16 March 2005 (16.03.2005) US

(71) Applicant *(for all designated States except US)*: **SIDNEY KIMMEL CANCER CENTER** [US/US]; 1085 Road to the Cure, San Diego, CA 92121 (US).

(72) Inventor; and

(75) Inventor/Applicant *(for US only)*: **GLINKSII, Gennadi, V.** [US/US]; 939 Coast Boulevard, Suite 4M, La Jolla, CA 92037 (US).

(74) Agents: **SEQUEIRA, Antonia, L.** et al.; FENWICK & WEST, LLP, 801 California Street, Mountain View, CA 94041 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, LY, MA, MD, MG, MK, MN, MW, MX, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SM, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

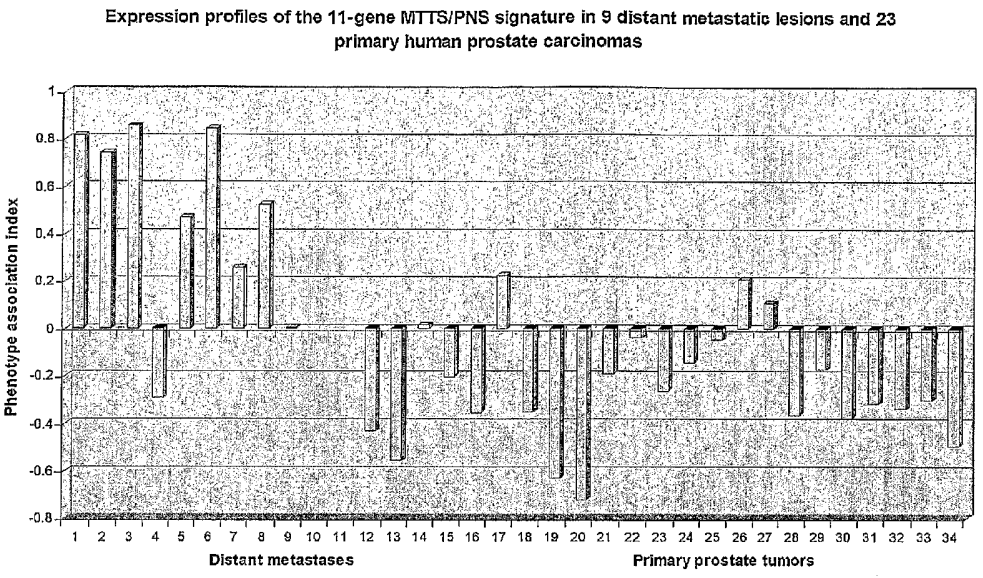
(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IS, IT, LT, LU, LV, MC, NL, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: METHODS AND COMPOSITIONS FOR PREDICTING DEATH FROM CANCER AND PROSTATE CANCER SURVIVAL USING GENE EXPRESSION SIGNATURES



(57) Abstract: The emerging concept of cancer stem cells suggests that activation in transformed cells of "sternness" genetic pathways (e.g., normal stem cells' self-renewal pathways) may contribute to the survival life cycle of cancer stem cells, and to tumor progression and metastasis of the malignancy. Thus, activation of "sternness" genes in cancer cells may be associated with aggressive clinical behavior and increased likelihood of therapy failure. General methods and kits associated with prediction of clinical outcome for a disease state of a subject based on gene expression analysis are described. The invention includes determining expression of at least three genes selected from the group consisting of GBX2, MKI67, CCNB1, BUB1, KNTC2, USP22, HCFC1, RNF2, ANK3, FGFR2, and CES1, and mouse homologs thereof.

WO 2006/110264 A2

TITLE

[001] METHODS AND COMPOSITIONS FOR PREDICTING DEATH FROM CANCER AND PROSTATE CANCER SURVIVAL USING GENE EXPRESSION SIGNATURES

CROSS REFERENCE TO RELATED APPLICATIONS

[002] This application claims the benefit of U.S. Provisional Application No. 60/663,014, filed March 16, 2005, which is herein incorporated by reference in its entirety for all purposes.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH OR DEVELOPMENT

[003] The U.S. Government has certain rights in this invention pursuant to Grant No. 5RO1 CA89827 awarded by the National Institutes of Health (National Cancer Institute).

FIELD OF THE INVENTION

[004] The present invention relates to predicting clinical outcome of patients by detecting gene expression patterns relating to molecular signatures.

BACKGROUND OF THE INVENTION

[005] Studies regarding the genetic basis of human cancer progression have allowed many advances toward finding effective treatments for this disease. Beyond providing an effective treatment for cancer, genetic analyses can provide other essential information about progression of the disease. Cancer patients in the early stages of the disease, for example, would typically greatly benefit from simply knowing more about the aggressiveness that their cancer is likely to exhibit, how their cancer is likely to progress, whether it is likely to metastasize, whether it is likely to recur after therapy (and how quickly it might recur), and so forth. With this type of knowledge in hand, physicians could respond by applying more aggressive therapies for patients with cancers that will likely exhibit particularly aggressive malignant behavior. Treatments could be properly tailored to the patient based on prognosis for that patient's particular disease state.

[006] Recent studies suggest that more aggressive cancers may have some recognizable and measurable characteristics that distinguish them from the less aggressive types. Studies suggest that some types of cancers include a small number of cells in tumors with significant biological resemblance to stem cells, which are unspecialized, precursor cells with the ability to quickly divide and differentiate to give rise to specific specialized cells (Al-Hajj, M., Wicha, M.S., et al., M.F. Prospective identification of tumorigenic breast cancer cells. Proc. Natl. Acad. Sci. USA 2003, 100:3983-3988; Pardal, R., Clarke, M.F., Morrison, S.J. Applying the principle of stem-cell biology to cancer. Nature Review Cancer 2003, 3:895-902; Smalley, M. and Ashworth, A. Stem cells and breast cancer: a field in transit. Nature Review Cancer 2003, 3:832-844, each incorporated herein by reference). For a pluripotent stem cell-like phenotype, self-renewal ability is an essential defining property distinguishing stem cells from other cell types (Dick, J.E. Self-renewal writ in blood. Nature 2003, 423:231-233, incorporated herein by reference). Similarly, in cancer stem cells, this self-renewal ability can play an important role in tumor development, especially in more aggressive cancers. This small population of cancer stem cells within tumors can allow replication that seeds the growth of additional cancer cells. The presence of a rare stem-cell resembling population of cancer cells among the heterogeneous mix of cells comprising a tumor appears to be essential for sustained tumor growth and may contribute to the emergence of metastatic cancer cells during tumor progression (Pardal, R., Clarke, M.F.,

Morrison, S.J. Applying the principle of stem-cell biology to cancer. Nature Review Cancer 2003, 3:895-902; Al-Hajj, M., et al., Prospective identification of tumorigenic breast cancer cells. Proc. Natl. Acad. Sci. USA 2003, 100:3983-3988; Smalley, M. and Ashworth, A. Stem cells and breast cancer: a field in transit. Nature Review Cancer 2003, 3:832-844, incorporated herein by reference).

[007] This concept of cancer stem cells further implies that common genetic pathways might define critical stem cell-like functions in neoplastic stem cells, as well as in normal stem cells (Lessard, J. and Sauvageau, G. *BMI-1* determines the proliferative capacity of normal and leukaemic stem cells. Nature 2003, 423:255-260; Pardal, R., Clarke, M.F., Morrison, S.J. Applying the principle of stem-cell biology to cancer. Nature Review Cancer 2003, 3:895-902, incorporated herein by reference). In colorectal cancer, for example, constitutive activation of the β-catenin/TCF-4 pathway imposes a crypt progenitor phenotype on colorectal cancer cells, suggesting that analysis of normal stem cells and cancer cells may reveal common stem cell-like pathways engaged in malignant cells (van den Wetering, M., Sancho, E., Verweij, C., et al. The β-catenin/TCF-4 complex imposes a crypt progenitor phenotype on colorectal cancer cells. *Cell* 2002, 111:241-250, incorporated herein by reference).

[008] Specifically, genes associated with the potential of a stem cell to proliferate are likely to be of particular interest in cancer studies. As one example, recent studies indicate that the *Polycomb* group (PcG) gene *BMI-1* determines the proliferative potential of normal and leukemic stem cells and is required for the self-renewal of hematopoietic and neural stem cells (Lessard, J. and Sauvageau, G. *BMI-1* determines the proliferative capacity of normal and leukaemic stem cells. Nature 2003, 423:255-260; Park, I.-K., et al., *BMI-1* is required for maintenance of adult self-renewing haematopoietic stem cells. Nature 2003, 423:302-305; Molofsky, A.V., et al., *BMI-1* dependence distinguishes neural stem cell self-renewal from progenitor proliferation. Nature 2003, 425:962-967, each incorporated herein by reference). *BMI-1* oncogene is expressed in all primary myeloid leukemia and leukemic cell lines that have been analyzed in various studies so far and over-expression of *BMI-1* causes neoplastic transformation of lymphocytes (Lessard, J. and Sauvageau, G. *BMI-1* determines the proliferative capacity of normal and leukaemic stem cells. Nature 2003, 423:255-260; Lessard, J., et al., Stage-specific expression of polycomb group genes in human bone marrow cells. Blood 1998, 91:1216-1224; Haupt, Y., et al., J.M. *BMI-1* transgene induces lymphomas and collaborates with Myc in tumorigenesis. Oncogene 1993, 8:3161-3164; Alkema, M.J., et al., A. Perturbation of B and T cell development and predisposition to lymphomagenesis in Eμ-*BMI-1* transgenic mice require the *BMI-1* RING finger. Oncogene 1997, 15:899-910, each incorporated herein by reference). Recently, *BMI-1* expression was reported in human non-small-cell lung cancer and breast cancer cell lines, suggesting an oncogenic role for *BMI-1* activation in epithelial malignancies (Vonlanthen, S., et al. The *BMI-1* oncoprotein is differentially expressed in non-small-cell lung cancer and correlates with INK4A-ARF locus expression. Br. J. Cancer 2001, 84:1372-1376; Dimri, G.P., et al., The *BMI-1* oncogene induces telomerase activity and immortalizes human mammary epithelial cells. Cancer Res. 2002, 62:4736-4745; LaTulippe, E., et al., Comprehensive gene expression analysis of prostate cancer reveals distinct transcriptional programs associated with metastasis. Cancer Res. 2002, 62:4499-4506, each incorporated herein by reference).

[009] These strong ties between neoplastic stem cells and normal stem cells, and the common genetic pathways defining critical stem cell-like functions in cancer cells, provide a useful opportunity for further analysis. Expression profiling of tumor samples using oligonucleotide or cDNA microarray technology is a powerful tool for revealing multiple gene expression signatures associated with various cancers. For example, comparative gene expression profiling analysis of normal stem cells and cancer cells may reveal gene expression signatures of "stemness" pathways engaged in malignant cells. These gene signatures identified to be associated with certain cancers and identified to have an association with stem cell-like properties could then be used prognostically to predict clinical outcome for a particular patient. Accuracy of different technologies using expression profiling for providing diagnosis and prognosis could be increased through identification of small signatures that are highly effective in providing information regarding likely clinical outcome for a

cancer patient, even in the early stages of the cancer. These gene signatures could act as powerful predictors of distant metastasis, short interval to disease recurrence, death after therapy in cancer patients, and so forth, thus providing cancer patients with essential information before the cancer has had a chance to progress.

[010] Thus, there exists in the art a need for improved methods of predicting the clinical outcome of disease states, such as cancer, through use of gene signatures associated with genes that are differentially expressed or regulated in biological samples, such as tumor and normal cell samples. The present invention addresses these and other shortcomings of the art.

SUMMARY OF THE INVENTION

[011] Disclosed herein are kits and methods for predicting the clinical outcome for a disease state in a subject. Accordingly one aspect of the invention is a kit for predicting a clinical outcome for a disease state in a subject comprising a set of nucleic acid probes for determining expression level of a plurality of genes and instructions for use. The plurality of genes is selected from a group consisting of the genes of a gene set identified in Table 2 (described below). The set of nucleic acid probes is capable of hybridizing to RNA or cDNA species derived from the plurality of genes, and the probes allow quantification of the expression level and prediction of the clinical outcome based on said quantification.

[012] Another aspect is a method for predicting a clinical outcome for a disease state in a subject comprising detecting expression level of a plurality of genes in said subject. The plurality of genes is selected from a group consisting of the genes of a gene set identified in Table 2. A set of nucleic acid probes capable of hybridizing to RNA or cDNA species derived from the plurality of genes allows quantification of the expression level and prediction of the clinical outcome based on said quantification.

[013] In some embodiments of the kit and of the method, the plurality comprises all of the genes of the gene set identified in Table 2. In one embodiment, the plurality comprises the genes *MKI67* and *CCNB1*. In an embodiment where the disease state is prostate cancer, the plurality includes at least two genes selected from the group consisting of *MKI67*, *ANK3*, *FGFR2* and *CES1*. In an embodiment where the disease state is breast cancer, the plurality is selected from a group consisting of *CCNB1*, *BUB1*, and *KNTC2*. In still other embodiments, the plurality includes five or six of the genes identified in Table 2. In some embodiments, the invention further comprises analyzing a clinico-pathological feature selected from a group consisting of pre-RP Gleason sum, surgical margins, seminal vesicle invasion, age, and extra-capsular extension.

[014] In still another aspect of the invention, a kit is disclosed for predicting a clinical outcome for a disease state in a subject comprising a set of nucleic acid probes for determining expression level of a plurality of genes and instructions for use. The plurality of genes is selected from a group consisting of genes from gene set A identified in Table 9a, gene set B identified in Table 9b, gene set C identified in Table 9c, and gene set D identified in Table 9d (Tables described below). The set of nucleic acid probes is capable of hybridizing to RNA or cDNA species derived from the plurality of genes, and the probes allow quantification of the expression level and prediction of the clinical outcome based on said quantification. In certain embodiments, probes are directed to all genes from an identified gene set. In other embodiments, probes are directed to a subset of genes from an identified gene set.

[015] Another aspect is a method for predicting a clinical outcome for a disease state in a subject comprising detecting expression level of a plurality of genes in said subject. The plurality of genes is selected from a group consisting of genes from gene set A identified in Table 9a, gene set B identified in Table 9b, gene set C identified in Table 9c, and gene set D identified in Table 9d. A set of nucleic acid probes capable of hybridizing to RNA or cDNA species derived from the plurality of genes allows quantification of the expression level and prediction of the clinical outcome based on said quantification. In certain embodiments, probes are directed to all genes from an identified gene set. In other embodiments, probes are directed to a subset of genes from an identified gene set.

[016] In some embodiments of the methods, the genes are extracted from a tumor cell recovered from said subject. The tumor cell can be recovered from an organ selected from the group consisting of a prostate, a breast, a colon, a lung, a bladder, and an ovary.

[017] In some embodiments, the methods further comprise performing a Kaplan-Meier survival analysis to determine probability that the subject will remain disease-free for a time period after therapy. In some embodiments, the methods further comprise calculating a Pearson correlation coefficient by comparing an expression profile for a tumor sample taken from the subject to a stem cell-associated expression profile.

[018] In any one of the embodiments described above, the nucleic acid probes can be affixed to a solid support or the probes can comprise primers for nucleic acid amplification of a subset of genes. The primers can be selected from a group consisting of the primers identified in Table 5 and Table 6 (described below). Furthermore, in any of the embodiments described above, the disease state preferably is prostate cancer, breast cancer, lung cancer, ovarian cancer, bladder cancer, lymphoma, mantle cell lymphoma, mesothelioma, medulloblastoma, glioma, or acute myeloid leukemia. In addition, the prognosis can be selected from the group consisting of recurrence of the disease state after therapy, non-recurrence of the disease state after therapy, therapy failure, short interval to disease recurrence (e.g., less than two years, or less than one year, or less than six months), short interval to metastasis (e.g., less than two years, or less than one year, or less than six months), invasiveness, non-invasiveness, likelihood of metastasis, likelihood of distant metastasis, poor survival after therapy, death after therapy, and disease free survival.

[019] Another aspect of the present invention is a kit for determining expression of at least three genes selected from the group consisting of *GBX2*, *MKI67*, *CCNB1*, *BUB1*, *KNTC2*, *USP22*, *HCFC1*, *RNF2*, *ANK3*, *FGFR2*, and *CES1*, and mouse homologs thereof. The kit comprises a set of probes to specifically detect expression of the at least three genes and that specifically do not detect expression of other genes. The set of probes are nucleic acids or antibodies (the term "antibodies" can include antibodies, antibody fragments, scFvs, etc.).

[020] In some embodiments, the set of probes are nucleic acids capable of hybridizing under normal stringency conditions (e.g., conditions under which a compound of the invention will hybridize to its target sequence, but to a minimal number of other sequences, such as described in Korkola, et al., Optimizing Stringency for Expression Microarrays, Microarray Technologies 2003, 35:828-835 and in U.S. Patent No. 7,005,500, filed November 14, 2001, incorporated by reference) to RNA species transcribed from the at least three genes or to cDNA species derived from the RNA species. In some embodiments, the set of probes are PCR primers. Further, the PCR primers can be at least three pair of primers selected from the group consisting of SEQ. ID NO: 3, SEQ. ID NO: 4, SEQ. ID NO: 5, SEQ. ID NO: 6, SEQ. ID NO: 7, SEQ. ID NO: 8, SEQ. ID NO: 9, SEQ. ID NO: 10, SEQ. ID NO: 11, SEQ. ID NO: 12, SEQ. ID NO: 13, SEQ. ID NO: 14, SEQ. ID NO: 15, SEQ. ID NO: 16, SEQ. ID NO: 17, SEQ. ID NO: 18, SEQ. ID NO: 19, SEQ. ID NO: 20, SEQ. ID NO: 213, SEQ. ID NO: 22, SEQ. ID NO: 23, SEQ. ID NO: 24, SEQ. ID NO: 25, SEQ. ID NO: 26, SEQ. ID NO: 27, and SEQ. ID NO: 28.

[021] In some embodiments, the kit comprises a solid phase. Further, in some embodiments, the set of probes consists of at least three probe sets selected from the group consisting of Affymetrix HG-U95Av2 probe set 33688_at, Affymetrix HG-U95Av2 probe set 418_at, Affymetrix HG-U95Av2 probe set 34736_at, Affymetrix HG-U95Av2 probe set 41081_at, Affymetrix HG-U95Av2 probe set 40041_at, Affymetrix HG-U95Av2 probe set 39866_at, Affymetrix HG-U95Av2 probe set 37910_at, Affymetrix HG-U95Av2 probe set 33484_at, Affymetrix HG-U95Av2 probe set 36967_g_at, Affymetrix HG-U95Av2 probe set 1143_s_at, Affymetrix HG-U95Av2 probe set 37203_at, Affymetrix HG-U133A probe set 210560_at, Affymetrix HG-U133A probe set 212022_s_at, Affymetrix HG-U133A probe set 214710_s_at, Affymetrix HG-U133A probe set 216277_at, Affymetrix HG-U133A probe set 204162_at, Affymetrix HG-U133A probe set 216964_at, Affymetrix HG-U133A probe set 202473_x_at, Affymetrix HG-U133A probe set 205215_at, Affymetrix HG-U133A probe set 209442_x_at, Affymetrix HG-U133A probe set 208228_s_at, Affymetrix HG-U133A probe set 209616_s_at, Affymetrix

MG-U74A probe set 94200_at, Affymetrix MG-U74A probe set 99457_at, Affymetrix MG-U74A probe set 160159_at, Affymetrix MG-U74A probe set 104097_at, Affymetrix MG-U74A probe set 93441_at, Affymetrix MG-U74A probe set 97960_at, Affymetrix MG-U74A probe set 100901_at, Affymetrix MG-U74A probe set 93164_at, Affymetrix MG-U74A probe set 98477_s_at, Affymetrix MG-U74A probe set 93090_at, and Affymetrix MG-U74A probe set 101538_i_at.

[022] In some embodiments of the invention, the at least three genes are *CCNB1, BUB1, KNTC2,* or the mouse homologs thereof. In other embodiments, the kit is a kit for determining expression of *MKI67, ANK3, FGFR2,* and *CES1*, or the mouse homologs thereof, and the set of probes specifically detects expression of *MKI67, ANK3, FGFR2,* and *CES1,* or the mouse homologs thereof. In still other embodiments, the kit is a kit for determining expression of *GBX2, MKI67, CCNB1, BUB1, KNTC2, USP22, HCFC1, RNF2, ANK3, FGFR2,* and *CES1*, or the mouse homologs thereof, and the set of probes specifically detects expression of *GBX2, MKI67, CCNB1, BUB1, KNTC2, USP22, HCFC1, RNF2, ANK3, FGFR2,* and *CES1,* or the mouse homologs thereof.

[023] Another aspect of the present invention is a method for predicting a clinical outcome for a disease state in a subject. The method comprises obtaining a sample from said subject, and determining from the sample a set of gene expression measurements for at least three genes selected from the group consisting of *GBX2, MKI67, CCNB1, BUB1, KNTC2, USP22, HCFC1, RNF2, ANK3, FGFR2,* and *CES1*, or the mouse homologs thereof. The method further comprises determining a correlation coefficient between the set of gene expression measurements and a reference standard set of gene expression measurements obtained by comparing expression values from a stem cell and from a tumor cell for the set of genes. The sign of the correlation coefficient is predictive of the clinical outcome for the disease state.

[024] In some embodiments, the stem cell is a peripheral nervous system neurosphere. In some embodiments, the tumor cell is a metastatic prostate tumor cell. In addition, in some embodiments, the disease state is cancer, and in some embodiments, the cancer is prostate cancer. The cancer can also be selected from the group consisting of prostate cancer, breast cancer, lung cancer, ovarian cancer, bladder cancer, lymphoma, mantle cell lymphoma, mesothelioma, medulloblastoma, glioma, and acute myeloid leukemia. In some embodiments, the clinical outcome is selected from the group consisting of recurrence, therapy failure, likelihood of metastasis, likelihood of distant metastasis, disease free survival, invasiveness, and likelihood of survival at a predetermined time period.

[025] In some embodiments of the present invention, the at least three genes are *CCNB1, BUB1, KNTC2,* or the mouse homologs thereof. In other embodiments, the set of gene expression measurements are expression measurements of *MKI67, ANK3, FGFR2,* and *CES1*, or the mouse homologs thereof. In still other embodiments, the set of gene expression measurements are expression measurements of *GBX2, MKI67, CCNB1, BUB1, KNTC2, USP22, HCFC1, RNF2, ANK3, FGFR2,* and *CES1*, or the mouse homologs thereof.

[026] In some embodiments, the method further comprises analyzing a clinico-pathological feature selected from the group consisting of a pre-radical prostatectomy Gleason sum, a surgical margin evaluation, a seminal vesicle invasion, an age, and an extra-capsular extension.

BRIEF DESCRIPTION OF THE SEVERAL VIEWS OF THE DRAWING

[027] These and other features, aspects, and advantages of the present invention will become better understood with regard to the following description, and accompanying drawings, where:

[028] **Figure 1** is a graph showing microarray data-derived expression values of BMI-1 mRNA in multiple human prostate cancer cell lines established from metastatic tumors (PC-3, LNCap, DuCap, VCap, etc.) and normal human prostate epithelial cells, NPEC (NPEC, normal prostate epithelial cells).

[029] **Figure 2** is a graph showing an expression profile (depicted as a phenotype association index) of the 11-gene MTTS/PNS signature in metastatic lesions at multiple distant target organs and primary prostate carcinomas in the TRAMP transgenic mouse model of prostate cancer.

[030] **Figure 3** is a graph showing an expression profile (depicted as a phenotype association index) of the 11-gene MTTS/PNS signature in metastatic lesions at multiple distant target organs and primary prostate carcinomas in human prostate cancer patients.

[031] **Figure 4** is a graph showing Kaplan-Meier survival curves of prostate cancer patients with distinct expression profiles of the 11-gene MTTS/PNS signature.

[032] **Figure 5** is a graph showing Kaplan-Meier relapse-free survival curves of prostate cancer patients with distinct expression profile of the 11-gene MTTS/PNS signature. RP, radical prostatectomy.

[033] **Figure 6** is a graph showing the Kaplan-Meier survival curves for 79 prostate cancer patients stratified into distinct sub-groups using a weighted survival predictor score algorithm.

[034] **Figure 7** is a graph showing the Kaplan-Meier survival curves for distinct sub-groups of prostate cancer patients diagnosed with early stage disease (stages 1C and 2A).

[035] **Figure 8** is a graph showing Kaplan-Meier survival curves for 20 prostate cancer patients stratified into distinct sub-groups using Q-RT-PCR assay of the 11-gene signature

[036] **Figure 9** is a graph showing the Kaplan-Meier analysis of the probability that patients would remain metastasis-free or survive after therapy among 97 early stage breast cancer patients according to whether they had a good-prognosis or poor-prognosis signatures defined by the expression profiles of the 11-gene MTTS/PNS signature.

[037] **Figure 10** is a graph showing the Kaplan-Meier analysis of the probability that patients would remain metastasis-free or survive after therapy among 125 lung adenocarcinoma patients of all stages according to whether they had a good-prognosis or poor-prognosis signatures defined by the expression profiles of the 11-gene MTTS/PNS signature.

[038] **Figure 11** is a graph showing the Kaplan-Meier analysis of the probability that patients would remain metastasis-free or survive after therapy among 37 ovarian cancer patients of all stages according to whether they had a good-prognosis or poor-prognosis signatures defined by the expression profiles of the 11-gene MTTS/PNS signature.

[039] **Figure 12** is a graph showing the Kaplan-Meier analysis of the probability that patients would remain metastasis-free or survive after therapy among 31 bladder cancer patients according to whether they had a good-prognosis or poor-prognosis signatures defined by the expression profiles of the 11-gene MTTS/PNS signature.

[040] **Figure 13** is a graph showing Kaplan-Meier survival analysis of the probability of a therapy failure in cancer patients diagnosed with a non-epithelial cancer, lymphoma, and having distinct expression profiles of the 11-gene MTTS/PNS signature

[041] **Figure 14** is a graph showing the expression profile of the 23-gene "stemness" signature in primary prostate tumors from patients with recurrent disease resembling "stemness" transcript abundance patterns in highly metastatic PC3MLN4 orthotopic xenografts in nude mice.

[042] **Figure 15** is a graph showing the expression profile of the 16-gene "stemness" signature in primary prostate tumors from patients with recurrent disease resembling "stemness" transcript abundance patterns in distant prostate cancer metastases.

[043] **Figure 16** is a graph showing the expression profile of the 14-gene "stemness" signature in 8 recurrent versus 13 non-recurrent human prostate carcinomas.

[044] **Figure 17** is a graph showing the expression profile of the 5-gene "stemness" signature in primary prostate tumors from patients with recurrent disease resembling "stemness" transcript abundance patterns in highly metastatic PC3MLN4 orthotopic xenografts in nude mice.

[045] **Figure 18** is a graph showing the Kaplan-Meier analysis of the probability that patients would remain disease-free among 21 prostate cancer patients comprising a clinical outcome group 1 according to whether they had a good-prognosis or poor-prognosis signatures defined by the expression profiles of the 23-gene "stemness" signature.

[046] **Figure 19** is a graph showing the Kaplan-Meier analysis of the probability that patients would remain disease-free among 21 prostate cancer patients comprising a clinical outcome group 1 according to whether they had a good-prognosis or poor-prognosis signatures defined by the expression profiles of the 14-gene "stemness" signature.

[047] **Figure 20** is a graph showing the Kaplan-Meier analysis of the probability that patients would remain disease-free among 21 prostate cancer patients comprising a clinical outcome group 1 according to whether they had a good-prognosis or poor-prognosis signatures defined by the expression profiles of the 5-gene "stemness" signature.

[048] **Figure 21** is a graph showing the Kaplan-Meier analysis of the probability that patients would remain disease-free among 21 prostate cancer patients comprising a clinical outcome group 1 according to whether they had a good-prognosis or poor-prognosis signatures defined by the expression profiles of the 16-gene "stemness" signature.

[049] **Figure 22** is a graph showing the Kaplan-Meier analysis of the probability that patients would remain disease-free where patients had at least 2 positive signatures or at least 3 negative signatures.

[050] **Figure 23** is a graph showing the Kaplan-Meier analysis of the probability that patients would remain disease-free where patients had 4 positive signatures or 2 or 3 positive signatures, or 3 or 4 negative signatures.

[051] **Figure 24** is a graph showing the actual frequency of disease recurrence after radical prostatectomy in prostate cancer patients with distinct "stemness" gene expression profiles defined by the four "stemness" signature algorithm.

DETAILED DESCRIPTION OF THE INVENTION

Definitions

[052] All terms, unless specifically defined below, are intended to have their ordinary meanings as understood by those of skill in the art. Claimed masses and volumes are intended to encompass variations in the stated quantities compatible with the practice of the invention. Such variations are contemplated to be within, e.g. about $\pm$ 10 – 20 percent of the stated quantities. In case of conflict between the specific definitions contained in this section and the ordinary meanings as understood by those of skill in the art, the definitions supplied below are to control.

[053] "Differentially expressed" refers to the existence of a difference in the expression level of a gene as compared between two sample classes. Differences in the expression levels of "differentially expressed" genes preferably are statistically significant.

[054] "Tumor" is to be construed broadly to refer to any and all types of solid and diffuse malignant neoplasias including but not limited to sarcomas, carcinomas, leukemias, lymphomas, etc., and includes by way of example, but not limitation, tumors found within prostate, breast, colon, lung, and ovarian tissues.

[055] A "tumor cell line" refers to a transformed cell line derived from a tumor sample. Usually, a "tumor cell line" is capable of generating a tumor upon explant into an appropriate host. A "tumor cell line" line usually retains, in vitro, properties in common with the tumor from which it is derived, including, e.g., loss of differentiation, loss of contact inhibition, and will undergo essentially unlimited cell divisions in vitro.

[056] A "control cell line" refers to a non-transformed, usually primary culture of a normally differentiated cell type. In the practice of the invention, it is preferable to use a "control cell line" and a "tumor cell line" that are related with respect to

the tissue of origin, to improve the likelihood that observed gene expression differences are related to gene expression changes underlying the transformation from control cell to tumor.

[057] "Orthotopic" refers to the placement of cells in an organ or tissue of origin, and is intended to encompass placement within the same species or in a different species from which the cells are originally derived.

[058] The term "in vivo" refers to processes that occur in a living organism.

[059] It must be noted that, as used in the specification and the appended claims, the singular forms "a," "an" and "the" include plural referents unless the context clearly dictates otherwise.

Introduction

[060] Recently, a global gene expression profiling approach was successfully utilized to identify molecular signatures associated with activation of oncogenic pathways, targeted genetic manipulations, or cellular responses to physiological stimuli, and to build robust transcriptional identifiers reliably recognizing the engagement of corresponding pathways within the high complexity patterns of gene expression in experimental and clinical samples (Lamb, J., Ramaswamy, S., et al., A mechanism of cyclin D1 action encoded in the patterns of gene expression in human cancer. Cell 2003, 114:323-334; Chang, H.Y., et al., Gene expression signature of fibroblast serum response predicts human cancer progression: Similarities between tumors and wounds. PLOS Biology 2004, 2:1-9; Raaphorst, F.M. et al., Poorly differentiated breast carcinoma is associated with increased expression of the human polycomb group EZH2 gene. Neoplasia 2003, 5:481-488, each incorporated herein by reference). The present invention uses techniques, such as microarray gene expression analysis, to determine whether invasive tumors, while actively seeding metastatic cancer cells as well as established distant metastatic lesions, have gene expression profiles similar to the transcriptional program of stem cells. This gene expression profiling approach was successfully utilized to identify molecular signatures associated with activation of oncogenic pathways and which consistently displayed a stem-cell resembling profile in distant metastatic lesions. Analyses of metastases and primary tumors from a transgenic mouse model of prostate cancer and from human cancer patients were conducted. The methods of the present invention were then used to estimate the prognostic power of the identified "stemness" signatures in predicting the clinical outcome for a cancer patient.

[061] In some embodiments of the present invention, in identifying stem cell-like signatures that can be used in predicting clinical outcome (as applied to the analysis of tumor samples), gene expression data showing genes up-regulated or down-regulated in primary tumors and metastases is compared to data showing genes up- or down-regulated in certain stem cells (e.g., in neural stem cells, hematopoeitic stem cells, embryonic stem cells, etc.). Sets of differentially regulated transcripts can be identified for distant metastatic lesions and primary tumors versus the stem cell samples. One or more genes are selected that have met the screening criterion requiring that the genes be differentially expressed between tumor and control cell lines or between tumor and normal clinical samples. Molecular signatures can then be identified from these sets of transcripts exhibiting concordant expression changes between metastatic tumor and stem cell samples. A more detailed explanation of methods that can be used to identify and validate the outcome prediction capabilities of these signatures is provided in Glinsky, Gennadi V. et al, Microarray analysis identifies a death-from-cancer signature predicting therapy failure in patients with multiple types of cancer, J. Clin. Invest. 2005, 1:115(6):1503-1521 (incorporated by reference), and in pending U.S. Patent Application Serial Number 10/861,003, filed June 3, 2004 and pending U.S. Patent Application Serial Number 10/660,434, filed September 10, 2003, each of which is incorporated herein by reference in its entirety.

[062] The molecular signatures can be used to predict the clinical outcome of a disease state (such as cancer) for patients. Although most of the description contained herein focuses primarily on prediction of clinical outcomes associated with cancer, the present invention can also be used for predicting clinical outcomes associated with other disease states (e.g., atherosclerosis, arthritis, etc.).

[063] In a broad and general sense, as applied to the analysis of tumor samples, the method of the present invention includes specifically detecting the expression level of a plurality of genes in a patient, where the genes correspond to one or more gene signatures identified using the procedures described above. Examples of specific signatures identified include those shown in Tables 2, 9a, 9b, 9c, and 9d, described in a later section. The molecular signatures identified can vary in the number of interrogated genes. In some embodiments, the molecular signature used includes at least 5, 11, 14, 16, 23 genes, or other number of genes that is found to be effective as a set in predicting clinical outcome. In some embodiments, one or more of the genes contained in the gene set for each molecular signature is used for predicting clinical outcome for a patient. In some embodiments, at least two or more of the signatures identified in the Tables 2, 9a, 9b, 9c, or 9d are used in the methods or in a kit of the present invention to predict clinical outcome for a patient.

[064] Specifically detecting expression would be understood by one of skill in art, in case of a nucleic acid probe, to include measuring the level of mRNA or a cDNA to which a probe has been engineered to bind, where the probe binds the intended species and provides a distinguishable signal. Exemplary methods for selecting PCR primers and/or hybridization probes are included in Innis et al., eds., 1990, PCR Protocols: A Guide to Methods and Applications, Academic Press Inc., San Diego, Calif; Froehler et al., 1986, Nucleic Acid Res. 14:5399–5407; McBride et al., 1983, Tetrahedron Lett. 24:246–248, U.S. Patent No. 7,013221, filed April 28, 2000, incorporated by reference. Preferably probes have length of at least 20 nucleotides which provides requisite specificity for detecting expression, although they may be shorter depending upon other species expected to be found in sample. Specifically detecting expression for measurement or determing protein expression levels can also be accomplished by using a specific binding reagent, such as an antibody, as described in more detail below.

[065] In some embodiments, the kits and methods of the present invention can be used to predict various different types of clinical outcomes. For example, the invention can be used to predict recurrence of a disease state after therapy, non-recurrence of a disease state after therapy, therapy failure, short interval to disease recurrence, short interval to metastasis in cancer, invasiveness, non-invasiveness, likelihood of metastasis in cancer, likelihood of distant metastasis in cancer, poor survival after therapy, death after therapy, disease free survival, and so forth.

[066] In some embodiments, a set of nucleic acid probes capable of hybridizing to RNA or cDNA species derived from the plurality of genes making up the molecular signature allows quantification of the expression level and prediction of the clinical outcome based on this quantification. In some embodiments, the probes are affixed to a solid support, such as a microarray (such as those provided by Affymetrix at http://www.affymetrix.com). Methods for creating microarrays and examples of microarrays used the present invention are described in more detail below. In other embodiments, the probes are primers for nucleic acid amplification of set of genes. Methods for Q-RT-PCR used with the present invention are described in more detail below. In general, expression of the genes within the gene set of the molecular signature can be analyzed by any method now known or later developed to assess gene expression, including but not limited to measurements relating to the biological processes of nucleic acid amplification, transcription, RNA splicing, and translation. Thus, direct and indirect measures of gene copy number (e.g., as by fluorescence in situ hybridization or other type of quantitative hybridization measurement, or by quantitative PCR), transcript concentration (e.g., as by Northern blotting, expression array measurements, quantitative RT-PCR, or comparative genomic hybridization, CGH as described in e.g., U.S. Patent No. 6,335,167, incorporated by reference), and protein concentration (e.g., by quantitative 2-D gel electrophoresis, mass spectrometry, Western blotting, ELISA, or other method for determining protein concentration).

[067] One of ordinary skill in the art would recognize that different affinity reagents could be used with present invention, such as one or more antibodies (e.g., monoclonal or polyclonal antibodies) and the invention can include using techniques, such as ELISA, for the analysis. Thus, specific antibodies (e.g., specific to the genes of the proteins encoded by the molecular signature of interest) can be used in a kit and in methods of the present invention for predicting clinical outcome based on expression analysis in a manner similar to the kits and methods described above. In the case of antibodies and related affinity reagents such as, e.g., antibody fragments, and engineered sequences such as single chain Fvs (scFvs),

these reagents must specifically bind their intended target, i.e., a protein encoded by a gene included in the molecular signature of interest. Specific binding includes binding primarily or exclusively to an intended target. Specific binding is easily assessed using, e.g., a Western blot, where the reagent gives rise to a band at the expected molecular weight that is at least 2 or at least 10 or more times intense than other bands that might appear on the gel. For example, in a kit of this embodiment, the kit would include reagents and instructions for use, where the reagents are antibodies and the antibodies hybridize to the plurality of expression products of the gene set consisting of genes identified in Table 3 or the antibodies hybridize to the plurality of expression products selected from a group consisting of genes from gene set A identified in Table 9a, gene set B identified in Table 9b, gene set C identified in Table 9c, gene set D identified in Table 9d. It is well-known in the art the manner in which antibodies can be identified and obtained from a variety of sources, such as the MSRS catalog of antibodies (Aerie Corporation, Birmingham, MI), or can be prepared via conventional antibody-generation methods. Methods for preparation of polyclonal antisera are taught in, for example, Ausubel, F.M. et al., Current Protocols in Molecular Biology, Volume 2, 1997, pp. 11.12.1-11.12.9 (incorporated by reference). Preparation of monoclonal antibodies is taught in, for example, Ausubel, F.M. et al., Current Protocols in Molecular Biology, Volume 2, 1997, pp. 11.4.1-11.11.5 (incorporated by reference). Preparation of scFvs is taught in, e.g., U.S. Patent Nos. 5,516,637 and 5,872,215, both of which are incorporated by reference.

[068] Signatures identified (such as those exhibiting the most significant correlation of expression profiles in stem cells and cancer metastasis) can be used to discriminate between metastatic and primary prostate tumors in patients, and thus can be used in predicting clinical outcome for patients. In some embodiments, a survival prediction model based on a signature is validated by testing the prognostic performance of the model in multiple independent therapy outcome data sets representing disease states (e.g., epithelial and non-epithelial cancers). A prognosis discrimination cut-off value for a signature can be selected based on highest level of statistical significance in patient's stratification into poor and good prognosis groups as determined by a log-rank test (lowest P value and highest hazard ratio).

[069] In some embodiments, to assess a potential diagnostic and prognostic relevance of the signatures, a Pearson correlation coefficient is calculated (e.g., using Microsoft Excel and the GraphPad Prism version 4.00 software) for each individual tumor sample by comparing the expression profiles of individual samples to the reference expression profile in stem cells. The Pearson correlation coefficient can be used to measure degree of resemblance of the transcript abundance rank order within a gene cluster between a sample and reference standard, which can be designated as a phenotype association index (PAI). Samples with stem cell-resembling expression profiles (stem cell-like PAI or SPAI) are expected to have positive values of Pearson correlation coefficients. Clinical samples with the Pearson correlation coefficient at or higher than the cut-off value can be identified as having the poor prognosis signature. Clinical samples with the coefficient lower than the cut-off value were identified as having the good prognosis signature. In some embodiments, the survival prediction model performance is confirmed using sample stratification approaches, such as terrain clustering, support vector machine classification, and weighted survival score algorithm.

[070] In some embodiments, the potential clinical utility of a signature can be further validated by evaluating the prognostic power of the signature applied to samples obtained from cancer patients who developed recurrence after therapy and to other patients who remained disease-free. A Kaplan-Meier survival analysis can be used to determine if there is a highly significant difference in the probability that cancer patients would remain disease-free after therapy between groups with positive and negative SPAIs defined by the signature. An estimated hazard ratio for disease recurrence after therapy can be determined for patients with positive versus negative SPAIs defined by the signature.

[071] In some embodiments, to ascertain the incremental statistical power of the individual covariates as predictors of therapy outcome and unfavorable prognosis, univariate and multivariate Cox proportional hazard survival analyses are performed. These analyses allow comparison of the prognostic performance of an entire stemness signature and of individual genes making up the signature or subsets of genes.

[072] In some embodiments, a weighted survival score analysis is implemented to reflect the incremental statistical power of the individual covariates as predictors of therapy outcome based on a multi-component prognostic model. Final survival predictor score can comprise a sum of scores for individual genes of a signature and can reflect the relative contribution of each gene in the multivariate analysis. The negative weighting values imply that higher expression correlates with longer survival and favorable prognosis, whereas positive scores indicate that higher expression correlates with poor outcome and shorter survival. Application of this weighted survival predictor model based on cumulative score of weighted expression values of genes making up a signature can be used to confirm the prognostic power of the identified signature in stratification of cancer patients into sub-groups with statistically distinct probability of relapse-free survival after therapy.

[073] Similar types of methods (e.g., Kaplan-Meier methods) can also be used to determine a signature's prediction capabilities of a short relapse survival after therapy in patients with an early stage disease, of metastatic recurrence, and of poor survival after therapy. In addition, Kaplan-Meier analysis can be used to determine the probability of developing distant metastases after therapy and higher risk of death after therapy. These analyses can be used to examine the predictive capabilities of signatures regarding numerous types of cancer, both epithelial and non-epithelial. Further detail regarding the Kaplan-Meier analysis and other methods is provided in Glinsky, Gennadi V. et al, Microarray analysis identifies a death-from-cancer signature predicting therapy failure in patients with multiple types of cancer, J. Clin. Invest. 2005, 1:115(6):1503-1521 (incorporated by reference).

[074] More detailed information regarding the methods/kits of the present invention and how these methods are applied for detecting expression, including methods and kits involving an 11-gene signature in the first example and four other stemness signatures in the second example, is included below.

EXAMPLES

[075] Below are examples of specific embodiments for carrying out the present invention. The examples are offered for illustrative purposes only, and are not intended to limit the scope of the present invention in any way. Efforts have been made to ensure accuracy with respect to numbers used (e.g., amounts, temperatures, etc.), but some experimental error and deviation should, of course, be allowed for.

[076] The practice of the present invention will employ, unless otherwise indicated, conventional methods of protein chemistry, biochemistry, recombinant DNA techniques and pharmacology, within the skill of the art. Such techniques are explained fully in the literature. See, e.g., T.E. Creighton, *Proteins: Structures and Molecular Properties* (W.H. Freeman and Company, 1993); A.L. Lehninger, *Biochemistry* (Worth Publishers, Inc., current addition); Sambrook, et al., *Molecular Cloning: A Laboratory Manual* (2nd Edition, 1989); *Methods In Enzymology* (S. Colowick and N. Kaplan eds., Academic Press, Inc.); *Remington's Pharmaceutical Sciences*, 18th Edition (Easton, Pennsylvania: Mack Publishing Company, 1990); Carey and Sundberg *Advanced Organic Chemistry 3rd Ed.* (Plenum Press) Vols A and B(1992).

Materials and Methods

[077] The materials and methods used with regard to the present invention are described in detail in Glinsky, Gennadi V. et al, Microarray analysis identifies a death-from-cancer signature predicting therapy failure in patients with multiple types of cancer, J. Clin. Invest. 2005, 1:115(6):1503-1521 (incorporated by reference), and some of the methods are also described in pending U.S. Patent Application Serial Number 10/861,003, filed June 3, 2004 and pending U.S. Patent Application Serial Number 10/660,434, filed September 10, 2003, each of which is incorporated herein by reference in its entirety. Specifically, the incorporated references describe the materials and methods associated with the use of clinical samples and cell cultures, anoikis assay, apotosis assay for identifying and quantifying apoptotic cells, use of flow cytometry, development of orthotopic xenografts of human prostate PC-3 cells and sublines, creation of the transgenic

mouse model of prostate cancer, tissue processing for mRNA and RNA isolation, RNA and mRNA extraction, usage of Affymetrix arrays for mRNA quality control and gene expression analysis, and data analysis..

[078] The detailed protocol of discovery of an 11-gene signature associated with the *BMI-1* pathway in stem cells, including the steps for identification of differentially regulated transcripts in the TRAMP mouse model, PNS (peripheral nervous system) neurospheres, and CNS (central nervous system) neurospheres, identification of sub-sets of transcripts exhibiting concordant expression changes, selection of small gene clusters from the sub-sets (e.g., to obtain the 11-gene MTTS (metastatic TRAMP tumor sample)/PNS signature, the 11-gene MTTS/CNS signature, and the 14-gene MTTS/PNS/CNS signature), testing the three signatures for metastatic phenotype-discriminative power leading to selection of the best-performing 11-gene MTTS/PNS signature (also referred to as 11-gene signature or 11-gene *BMI-1* pathway signature) for further validation analysis, are described in detail in Glinsky, Gennadi V. et al, Microarray analysis identifies a death-from-cancer signature predicting therapy failure in patients with multiple types of cancer, use of the SPAI Index, Cox analysis, random co-occurrence test, J. Clin. Invest. 2005, 1:115(6):1503-1521 (incorporated by reference). In addition, these methods are described with regard to the Examples below.

Validation of the 11-Gene Signature

SPAI Index

[079] Definition of the Pearson correlation coefficient as a phenotype association index [stem cell-resembling phenotype association indices (SPAIs)] is based on highly concordant behavior of the 11-gene signature between neural stem cells in the state of PNS neurospheres and prostate cancer metastasis ($r = 0.9897$; $P < 0.0001$). A standard PNS neurosphere and TRAMP metastasis values were established as described in the signature discovery protocol. They were used as uniform reference standards for measurements of Pearson correlation coefficients for clinical samples consistently throughout the study.

[080] A degree of resemblance of the transcript abundance rank order within a gene cluster between a test sample and reference standard is measured by a Pearson correlation coefficient and designated as a phenotype association index (PAI). Samples with stem cell-resembling expression profiles (stem cell-like PAI or SPAI) are expected to have positive values of Pearson correlation coefficients.

Random co-occurrence test.

[081] We performed 10,000 permutations test to check how likely small 11-gene signatures derived from the large MTTS signature would display high discrimination power to assess the significance at the 0.1% level. We carried out 10,000 permutations of small 11-gene signatures derived from the large 1345-gene MTTS signature and compared their sample stratification power to the 11-gene MTTS/PNS signature. The classification performance cut-off p-values were established by applying two-tailed T-test to the 11-gene MTTS/PNS signature ($p = 0.0005$ for metastasis versus primary prostate cancer data set and $p = 0.026$ for recurrent versus non-recurrent prostate cancer data set). Random concordant gene sets comprising ~200 transcripts were generated using mouse transcriptome data set representing expression profiling data of ~12,000 transcripts across 45 normal tissues (55). Inter- and intra-species array to array probe set match was performed at 95% or greater identity level using the Affymetrix data base (www.affymetrix.com).

[082] To assess discrimination of random 11-gene signatures derived from the 1345-gene MTTS signature two-tailed T-test was carried out for metastatic versus primary prostate cancer data set (32 samples) and recurrent versus non-recurrent prostate cancer data set (21 samples). The signatures were ranked based on p-values and ranking metrics of each random 11-gene signature were compared to the 11-gene MTTS/PNS signature p-values. We found that 10,000 permutations generated 7 random 11-gene signatures performing at sample classification level of the 11-gene MTTS/PNS signature.

Weighted survival predictor score algorithm

[083] We implemented the weighted survival score analysis to reflect the incremental statistical power of the individual covariates as predictors of therapy outcome based on a multi-component prognostic model. Microarray-based or Q-RT-PCR-derived gene expression values were normalized and log-transformed. The log-transformed normalized expression values for each data set were analyzed in a multivariate Cox proportional hazards regression model, with overall survival or event-free survival as the dependent variable.

[084] To calculate the survival/prognosis predictor score for each patient, we multiplied the log-transformed normalized gene expression value measured for each gene by a coefficient derived from the multivariate Cox proportional hazard regression analysis. The final survival predictor score comprises a sum of scores for individual genes and reflects the relative contribution of each of the eleven genes in the multivariate analysis. Negative weighting values indicate that higher expression correlates with longer survival and favorable prognosis, whereas positive weighting values indicate that higher expression correlates with poor outcome and shorter survival. Thus, the weighted survival predictor model is based on a cumulative score of the weighted expression values of eleven genes. Target siRNA SMART pools for BMI-1 and control luciferase siRNAs were purchased from Dharmacon Research, Inc. They were transfected into PC-3-32 human prostate carcinoma cells according to the manufacturer's protocols. Cell cultures were continuously monitored for growth and viability and assayed for mRNA expression levels of BMI-1 and selected set of genes using RT-PCR and Q-RT-PCR methods.

Quantitative RT-PCR analysis

[085] Real time PCR methods measure the accumulation of PCR products by a fluorescence detector system and allow for quantification of the amount of amplified PCR products in the log phase of the reaction. Total RNA was extracted using RNeasy mini-kit (Qiagen, Valencia, CA, USA) following the manufacturer's instructions. A measure of 1 μg (tumor samples), or 2 μg and 4 μg (independent preparations of reference cDNA samples) of total RNA was used then as a template for cDNA synthesis with SuperScript II (Invitrogen, Carlsbad, CA, USA). QPCR primer sequences were selected for each cDNA with the aid of Primer Express™ software (Applied Biosystems, Foster City, CA, USA). PCR amplification was performed with the gene-specific primers listed in Tables 5 and 6 (described in detail below).

[086] Q-PCR reactions and measurements were performed with the SYBR-Green and ROX as a passive reference, using the ABI 7900 HT Sequence Detection System (Applied Biosystems, Foster City, CA, USA). Conditions for the PCR were as follows: one cycle of 10 min at 95°C; 40 cycles of 0.20 min at 94° C; 0.20 min at 60°C and 0.30 min at 72°C. The results were normalized to the relative amount of expression of an endogenous control gene GAPDH.

[087] Expression of messenger RNA (mRNA) for eleven genes and an endogenous control gene (GAPDH) was measured in twenty specimens of primary prostate cancer obtained from patients with documented PSA recurrence within five years after RP (radial prostatectomy) and patients who remained disease-free for at least five years after RP (ten patients in each group) by real-time PCR method on an ABI PRISM 7900 HT Sequence Detection System (Applied Biosystems). For each gene at least two sets of primers were tested and the set-up with highest amplification efficiency was selected for the assay used in this study. Specificity of the assay for mRNA measurements was confirmed by the absence of the expected PCR products when genomic DNA was used as a template. Glyceraldehyde-3-phosphate dehydrogenase (GAPDH: 5'-CCCTCAACGACCACTTTGTCA-3' (SEQ ID NO: 1) and 5'- TTCCTCTTGTGCTCTTGCTGG- 3' (SEQ ID NO: 2)) was used as the endogenous RNA and cDNA quantity normalization control. For calibration and generation of standard curves, we used several reference cDNAs: cDNA prepared from primary in vitro cultures of normal human prostate epithelial cells (NPEC), cDNA derived from the PC-3M human prostate carcinoma cell line, and cDNA prepared from normal human prostate (NHP) (Glinsky, G.V., et al., Microarray analysis of xenograft-derived cancer cell lines representing multiple experimental models of human prostate cancer. Molecular Carcinogenesis 200337:209-221 (Magee, J.A., et al., Expression profiling reveals hepsin overexpression in prostate cancer. Cancer Res. 2001, 61:5692-5696, incorporated by reference).

[088] Expression analysis of all genes was assessed in two independent experiments using reference cDNAs to control for variations among different Q-RT-PCR experiments. Prior to statistical analysis, the normalized gene expression values were log-transformed similarly to the transformation of the array-based gene expression data.

Survival analysis

[089] Kaplan-Meier survival analysis was carried out using GraphPad Prism version 4.00 software (GraphPad Software, San Diego, CA; http:// www.graphpad.com). The end point for survival analysis in prostate cancer was the biochemical recurrence defined by serum PSA increase after therapy. Disease-free interval (DFI) was defined as the time period between the date of radical prostatectomy (RP) and the date of PSA relapse (recurrence group) or date of last follow-up (non--recurrence group). Statistical significance of the difference between the survival curves for different groups of patients was assessed using Chi square and Log-rank tests. To evaluate the incremental statistical power of the individual covariates as predictors of therapy outcome and unfavorable prognosis, we performed both univariate and multivariate Cox proportional hazard survival analyses.

Validation of Stemness Signatures in Predicting Clinical Outcome

Clinical Samples

[090] We utilized in our experiments three independent sets of human primary prostate tumors and distant metastases comprising 132 tissue samples. Microarray analysis and associated clinical information for 32 clinical samples (23 primary prostate tumors and 9 distant metastatic lesions) was utilized to delineate the expression profiles of human prostate cancer metastases were reported previously (11). Two clinical outcome sets comprising 21 (outcome set 1) and 79 (outcome set 2) samples were utilized for discovery and validation of the gene expression-based recurrence predictor algorithm. Original gene expression profiles of the 21 clinical samples (outcome set 1) analyzed in this study were reported elsewhere (Glinsky, G.V., et al.. Microarray analysis of xenograft-derived cancer cell lines representing multiple experimental models of human prostate cancer. *Molecular Carcinogenesis* 2003, 37:209-221, incorporated herein by reference). Primary gene expression data files of clinical samples as well as associated clinical information can be found at http://www-genome.wi.mit.edu/cancer/ . Further detail regarding clinical samples and cell cultures used can be found in Glinsky, Gennadi V. et al, Microarray analysis identifies a death-from-cancer signature predicting therapy failure in patients with multiple types of cancer, J. Clin. Invest. 2005, 1:115(6):1503-1521 (incorporated by reference).

Orthotopic Xenografts

[091] Orthotopic xenografts of human prostate PC-3 cells and sublines used in this study were developed by surgical orthotopic implantation as previously described (13). Briefly, 2×10^6 cultured PC3 cells, PC3M or PC3MLN4 sublines were injected subcutaneously into male athymic mice, and allowed to develop into firm palpable and visible tumors over the course of 2 - 4 weeks. Intact tissue was harvested from a single subcutaneous tumor and surgically implanted in the ventral lateral lobes of the prostate gland in a series of six athymic mice per cell line subtype. The mice were examined periodically for suprapubic masses, which appeared for all subline cell types, in the order PC3MLN4 >PC3M>>PC3. Tumor-bearing mice were sacrificed by CO2 inhalation over dry ice and necropsy was carried out in a $2-4^{o}$ C cold room. Typically, bilaterally symmetric prostate gland tumors in the shape of greatly distended prostate glands were apparent. Prostate tumor tissue was excised and snap frozen in liquid nitrogen. The elapsed time from sacrifice to snap freezing was < 5 min. A systematic gross and microscopic post mortem examination was carried out. Further detail regarding creation of the transgenic mouse model of prostate cancer, tissue processing for mRNA and RNA isolation, RNA and mRNA extraction, usage of Affymetrix arrays for mRNA quality control and gene expression, data analysis and survival analysis can be found in Glinsky, Gennadi V. et al, Microarray analysis identifies a death-from-cancer signature predicting therapy failure in patients with multiple types of cancer, J. Clin. Invest. 2005, 1:115(6):1503-1521 (incorporated by reference).

Data Analysis

[092] Detailed protocols for data analysis and documentation of the sensitivity, reproducibility and other aspects of the quantitative statistical microarray analysis using Affymetrix technology have been reported (Baron, V., et al., Inhibition of Egr-1 expression reverses transformation of prostate cancer cells in vitro and in vivo. *Oncogene* 2003, 22:4194-4204, incorporated by reference). 40-50% of the surveyed genes were called present by the Affymetrix Microarray Suite 5.0 software in these experiments. The concordance analysis of differential gene expression across the data sets was performed using Affymetrix MicroDB v. 3.0 and DMT v.3.0 software as described earlier (11, 13). We processed the microarray data using the Affymetrix Microarray Suite v.5.0 software and performed statistical analysis of expression data set using the Affymetrix MicroDB and Affymetrix DMT software. This analysis identified a set of 218 genes (91 up-regulated and 127 down-regulated transcripts) differentially regulated in tumors from patients with recurrent versus non-recurrent prostate cancer at the statistically significant level ($p<0.05$) defined by both T-test and Mann-Whitney test. The concordance analysis of differential gene expression across the clinical and experimental data sets was performed using Affymetrix MicroDB v. 3.0 and DMT v.3.0 software as described earlier. *See Id.* The Pearson correlation coefficient for individual test samples and appropriate reference standard was determined using the Microsoft Excel and the GraphPad Prism version 4.00 software. We calculated the significance of the overlap between the lists of "stemness" and prostate cancer-associated genes by using the hypergeometrical distribution tests.

EXAMPLE 1: 11-GENE SIGNATURE FOR PREDICTING CLINICAL OUTCOME IN PATIENTS

***BMI-1* oncogene expression is elevated in prostate cancer**

[093] Recent experimental observations documented an increased *BMI-1* expression in human non-small-cell lung cancer, human breast carcinomas, and established breast cancer cell lines, suggesting that an oncogenic role of the *BMI-1* activation may be extended beyond the leukemia and, perhaps, may affect progression of the epithelial malignancies as well (Vonlanthen, S., et al. The *BMI-1* oncoprotein is differentially expressed in non-small-cell lung cancer and correlates with INK4A-ARF locus expression. Br. J. Cancer 2001, 84:1372-1376; Dimri, G.P., et al. The *BMI-1* oncogene induces telomerase activity and immortalizes human mammary epithelial cells. Cancer Res. 2002, 62:4736-4745; LaTulippe, E., et al., Comprehensive gene expression analysis of prostate cancer reveals distinct transcriptional programs associated with metastasis. Cancer Res. 2002, 62:4499-4506; Gingrich, J.R., et al., Metastatic prostate cancer in a transgenic mouse. Cancer Res. 1996, 56:4096-4102). Microarray gene expression analysis of established cancer cell lines representing multiple experimental models of human prostate cancer revealed that *BMI-1* expression seems to be consistently elevated in human prostate cancer cell lines established from metastatic tumors (carcinoma cell lines used in this example were PC-3, DuCapL, DuCapR, Vcap, LNCap, PRO5, and LN3) compared to the primary cultures of human normal prostate epithelial cells (NPEC), as illustrated in FIG. 1 (Magee, J.A., et al., Expression profiling reveals hepsin overexpression in prostate cancer. Cancer Res. 2001, 61:5692-5696, incorporated by reference). To validate the results of the microarray experiments, quantitative reverse transcription-polymerase chain reaction (Q-RT-PCR) analysis of *BMI-1* mRNA expression was used, as shown in Table 1 below (showing the carcinoma cell lines for which expression was analyzed, and the average expression value, standard deviation, and P values for each).

Table 1. Q-RT-PCR analysis of the *BMI-1* mRNA expression in human prostate carcinoma cell lines

Cell line	Average Expression Value1	STDEV	P value2
NPEC	0.090656645	0.0154152	
LNCap	0.216610094	0.0311867	0.0013481
LNCapPro5	0.292913482	0.0222714	1.472E-05

LNCapLN3	0.235569094	0.0429103	0.0038571
PC-3	1.030811318	0.1271548	0.000586
PC-3LN4	0.635668126	0.0892679	0.0009314
PC-3Pro4	1.424229109	0.1758348	0.0005788
VCAP	0.192483261	0.012621	6.494E-05
DUCAP	0.128637764	0.012266	0.0092371

[1] Normalized average expression value from four measurements [2] Two-tailed T-test compared to the NPEC

Thus, results of expression profiling experiments appear to support the notion that transcriptional activation of the *BMI-1* gene is frequently associated with human prostate cancer.

[094] Interestingly, microarray analysis shows markedly higher *BMI-1* expression levels in lymph node metastases and highly metastatic orthotopic xenografts of human prostate carcinoma in nude mice compared to the less metastatic counterparts, implying that *BMI-1* activation might be associated with aggressive malignant behavior of prostate carcinoma cells. To test this hypothesis, expression profiling analysis of ~12,000 transcripts in a transgenic mouse model of metastatic prostate cancer was carried out. Microarray experiments detected increased levels of the *BMI-1* mRNA expression in late-stage invasive primary tumors and multiple distant metastatic lesions in the TRAMP transgenic mouse model of prostate cancer, thus, lending more credence to the idea linking the activation of *BMI-1*-associated pathway with prostate cancer metastasis.

Identification of a *BMI-1* pathway signature with concordant expression profiles in normal stem cells and distant metastatic lesions in a transgenic mouse model of prostate cancer

[095] Recent experiments established that the *BMI-1* gene is required for self-renewal of hematopoietic and neural stem cells and identified *BMI-1*-regulated genes in neural stem cells that are presumably engaged in an execution of self-renewal programs in a state of both central nervous system (CNS) and peripheral nervous system (PNS) neurospheres (Lessard, J. and Sauvageau, G. *BMI-1* determines the proliferative capacity of normal and leukaemic stem cells. Nature 2003, 423:255-260; Park, I.-K., et al., *BMI-1* is required for maintenance of adult self-renewing haematopoietic stem cells. Nature 2003, 423:302-305; Molofsky, A.V., et al., *BMI-1* dependence distinguishes neural stem cell self-renewal from progenitor proliferation. Nature 2003, 425:962-967, each incorporated herein by reference). It was hypothesized that molecular signatures associated with activation of a normal stem cells' self-renewal program in metastatic cancer cells might be possible to detect by looking for genes manifesting concordant patterns of regulation in metastasis and normal stem cells in $BMI\text{-}1^{+/+}$ versus $BMI\text{-}1^{-/-}$ genetic backgrounds. Therefore, a determination was made regarding whether expression profiles of transcripts activated and suppressed in prostate cancer metastases would recapitulate the expression profile of the *BMI-1* -regulated genes in normal stem cells by comparing the sets of differentially regulated genes in search for union/intersections of lists for both up- and down-regulated transcripts. This analysis identified genes exhibiting highly concordant profiles of transcript abundance behavior in prostate cancer metastases and $BMI\text{-}1^{+/+}$ versus $BMI\text{-}1^{-/-}$ PNS neurospheres, suggesting the presence of a conserved *BMI-1*-regulated pathway(s) similarly engaged in both normal stem cells and distant metastatic lesions of prostate carcinoma.

1) Identification of Parent Signatures

[096] Transgenic mouse models of prostate cancer (TRAMP) were used in these experiments. The metastatic TRAMP tumor samples (MTTS) signature is likely to be enriched for genes discriminative for the metastatic phenotype. It is

reasonable to assume that many of the gene expression patterns wired into the MTTS signature would manifest metastatic phenotype discriminative power and would have no relation to the transcriptional program of normal stem cells. These features of the MTTS signature were used for identification of the gene expression components of a stem cell transcriptome that are coordinately expressed in metastatic cancer cells and might manifest discriminative diagnostic power for the malignant phenotype. Sets of differentially regulated transcripts were independently identified for distant metastatic lesions and primary prostate tumors versus age-matched control samples in a transgenic TRAMP mouse model of metastatic prostate cancer (MTTS signature) as well as PNS (PNS signature) and CNS (CNS signature) neurospheres in *BMI-1*$^{+/+}$ versus *BMI-1*$^{-/-}$ backgrounds. This analytical step defined three large parent signatures: MTTS signature comprising 868 up-regulated and 477 down-regulated transcripts; PNS signature comprising 885 up-regulated and 1088 down-regulated transcripts; and CNS signature comprising 769 up-regulated and 778 down-regulated transcripts.

2) Identification of Concordant Sub-Sets of Genes (Child Signatures)

[097] The MTSS signature was intersected with the stem cell signatures in the state of PNS and CNS neurospheres to identify concordant sets of genes and define the stem cell signatures embedded into MTSS signature. Sub-sets of transcripts exhibiting concordant expression changes in metastatic TRAMP tumor samples (MTTS signature) as well as PNS (PNS signature) and CNS (CNS signature) neurospheres in *BMI-1*$^{+/+}$ versus *BMI-1*$^{-/-}$ backgrounds were identified. Thus, two concordant sub-sets of transcripts were identified corresponding to each binary comparison of metastatic TRAMP tumors and neural stem cell samples in a state of PNS and CNS neurospheres [141 up-regulated and 58 down-regulated transcripts for PNS neurospheres ($r = 0.7593$; $P < 0.0001$) and 40 up-regulated and 24 down-regulated for CNS neurospheres ($r = 0.7679$; $P < 0.0001$)]. A third concordant sub-set of 27 genes comprising 15 up-regulated and 12 down-regulated transcripts was selected for intersection common for all three signatures ($r = 0.8002$; $P < 0.0001$). Thus, three concordant sub-sets of genes were identified.

[098] This analysis also identified a stem cell-like expression profile for transcripts coordinately expressed in metastatic cancer cells and normal stem cells which can be used as a consistent reference standard to interrogate independent data sets for possible presence of a stem cell-like expression signature. From these concordant gene sets, we selected smaller gene expression signatures (e.g., 11 or 14 gene sets) comprising transcripts with high level of expression correlation in metastatic cancer cells and stem cells (the selection threshold for smaller signatures was arbitrarily set at Pearson correlation coefficients > 0.95). The reduction in the signature transcript number was terminated when further elimination of a transcript did not increase the value of the Pearson correlation coefficient. Using this approach a single candidate prognostic gene expression signature was selected for each binary intersection of the MTTS signature and parent stem cell signatures. The smaller child signatures (one 11-gene signature for the PNS set, one 11-gene signature for the CNS set, and one 14-gene signature for common PNS/CNS set) were tested for metastatic phenotype discriminative power and therapy outcome classification performance. As one example, the gene set for the 11-gene signature for the PNS set (the 11-gene MTTS/PNS signature) is shown below in Table 2.

Table 2. The 11-gene MTTS/PNS signature

GENE	UniGene (Homo sapiens)	Affymetrix HG-U95Av2 probe set	Affymetrix HG-U133A probe set	Affymetrix MG-U74A probe set	GenBank	Unigene (Mus Musculus)
GBX2	Hs.184945	33688_at	210560_at	94200_at	Z48800	Mm.204730
MKI67	Hs.80976	418_at	212022_s_at	99457_at	X82786	Mm.4078

CCNB1	Hs.23960	34736_at	214710_s_at	160159_at	X64713	Mm.379450
BUB1	Hs.469649	41081_at	216277_at	104097_at	AF002823	Mm.2185
KNTC2	Hs.414407	40041_at	204162_at	93441_at	AI595322	Mm.225956
USP22	Hs.462492	39866_at	216964_at	97960_at	AW125800	Mm.30602
HCFC1	Hs.83634	37910_at	202473_x_at	100901_at	U80821	Mm.248353
RNF2	Hs.124186	33484_at	205215_at	93164_at	Y12783	Mm.31512
ANK3	Hs.499725	36967_g_at	209442_x_at	98477_s_at	L40632	Mm.235960
FGFR2	Hs.533683	1143_s_at	208228_s_at	93090_at	M23362	Mm.16340
CES1	Hs.558865	37203_at	209616_s_at	101538_i_at	AW226939	Mm.22720

[099] Based on diagnostic and prognostic classification performance, a single best performing 11-gene MTTS/PNS signature was selected for further validation analysis. Based on the information provided in Table 2 above, one of ordinary skill in the art would recognize that further information about these genes is available from numerous sources, such as the National Center for Biotechnology at http://www.ncbi.nlm.nih.gov/ (e.g., by selecting "Gene" from the search window drop down menu for selection of databases to search and by conducting a search for the gene name (e.g., GBX2)). Exemplary cDNA and protein sequences for the genes shown in Table 2 are included in the Sequence Listing included herewith. In some embodiments the sequence used in the methods and kits of the invention comprises a sequence that has at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identity to the exemplified sequence included in the Sequence Listing.

[0100] The term percent "identity," in the context of two or more nucleic acid or polypeptide sequences, refer to two or more sequences or subsequences that have a specified percentage of nucleotides or amino acid residues that are the same, when compared and aligned for maximum correspondence, as measured using one of the sequence comparison algorithms described below (e.g., BLASTP and BLASTN or other algorithms available to persons of skill) or by visual inspection. Depending on the application, the percent "identity" can exist over a region of the sequence being compared, e.g., over a functional domain, or, alternatively, exist over the full length of the two sequences to be compared.

[0101] For sequence comparison, typically one sequence acts as a reference sequence to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are input into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. The sequence comparison algorithm then calculates the percent sequence identity for the test sequence(s) relative to the reference sequence, based on the designated program parameters.

[0102] Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith & Waterman, Adv. Appl. Math. 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, J. Mol. Biol. 48:443 (1970), by the search for similarity method of Pearson & Lipman, Proc. Nat'l. Acad. Sci. USA 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, Wis.), or by visual inspection (see generally Ausubel et al., infra).

[0103] One example of an algorithm that is suitable for determining percent sequence identity and sequence similarity is the BLAST algorithm, which is described in Altschul et al., J. Mol. Biol. 215:403-410 (1990). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (www.ncbi.nlm.nih.gov/).

3) Malignant Phenotype Classification Performance Tests

[0104] During the malignant phenotype classification performance tests, we asked whether individual metastatic lesions and primary prostate tumors would exhibit the stem cell-like expression profile of the candidate prognostic signatures. We selected for this analysis three small signatures demonstrating the most significant correlation of expression profiles in stem cells and prostate cancer metastasis. To assess a degree of similarity of the signature expression profiles in individual tumor samples and normal stem cells, we calculated a Pearson correlation coefficient for each sample by comparing signature expression profile in an individual sample to the stem cell-associated expression profile of the corresponding small signatures. Based on expected similarity of the prognostic signatures in stem cells and prostate cancer metastasis, we named the corresponding Pearson correlation coefficients measured for individual samples the stem cell-like phenotype association indices (SPAIs). As shown in FIG. 2, which illustrates the expression profile for one of the signatures, two of three late-stage invasive primary tumors and all distant metastatic lesions in the TRAMP transgenic mouse model of prostate cancer have positive SPAIs, thus, manifesting a stem cell-like expression profile of the small signatures.

Distant metastatic lesions and primary prostate tumors from cancer patients with differing therapy outcome display distinct expression profiles of the 11-gene MTTS/PNS signature

[0105] To perform similar analysis for human tumors, we translated the murine small signatures into list of human homologs using the Locuslink database (http://www.ncbi.nlm.nih.gov) and retrieved the expression data for corresponding Affymetrix probe sets. We calculated the SPAIs for each of 9 metastatic tumors and 23 primary prostate carcinomas and determined that seven of nine samples of distant metastatic lesions from prostate cancer patients exhibit a stem cell-like expression profile of the 11-gene MTTS/PNS signature, as illustrated in FIG. 3. In contrast, a majority of primary prostate tumors seem to display a distinct expression profile of the 11-gene MTTS/PNS signature as manifested in negative values of SPAIs). Interestingly, a sub-set of samples of primary prostate carcinomas manifests expression profiles of the 11-gene MTTS/PNS signature similar to the metastatic tumors as reflected in positive correlation coefficients (positive SPAI values in FIG. 3), suggesting that primary prostate tumors with distinct expression profiles of the PNS neurosphere-derived 11-gene MTTS/PNS signature (e.g., positive and negative values of SPAIs) may have different biological features and distinct clinical course of disease progression. Validation analysis using the CNS neurosphere-derived MTTS/CNS 11-gene signature and MTTS/PNS/CNS 14-gene signature indicates that application of these signatures is less informative in distinguishing metastatic and primary human prostate tumors in comparison to the MTTS/PNS 11-gene signature. Thus, we proceeded in our analyses with the MTTS/PNS 11-gene signature.

1) Evaluation of the clinical utility of the 11-Gene MTTS/PNS Signature

[0106] To evaluate the potential biological significance and clinical utility of the 11-gene MTTS/PNS signature expression in human prostate cancer, we set out to examine whether the detection of a stem cell-like expression profile in primary prostate tumors of individual cancer patients would help in patient's stratification at the time of diagnosis into sub-groups with distinct course of disease progression based on differing therapy outcome after radical prostatectomy. We assessed the prognostic power of the 11-gene MTTS/PNS signature based on ability to segregate the patients with recurrent and non-recurrent course of disease progression after radical prostatectomy into distinct sub-groups. We calculated a Pearson correlation coefficient for each of 21 tumor samples of outcome set 1 by comparing the 11-gene MTTS/PNS signature expression profiles of individual samples to the stem cell-like expression profile of the 11-gene MTTS/PNS signature in PNS neurospheres. To determine the prognostic power of the 11-gene MTTS/PNS signature, we performed Kaplan-Meier

survival analysis using as a clinical end-point disease-free interval (DFI) after therapy in prostate cancer patients with positive and negative SPAIs.

[0107] The Kaplan-Meier survival curves showed a highly significant difference in the probability that prostate cancer patients would remain disease-free after therapy between the groups with positive and negative SPAIs defined by the 11-gene MTTS/PNS signature, suggesting that patients with positive SPAIs exhibit a poor outcome signature whereas patients with negative SPAIs manifest a good outcome signature. As illustrated in FIG. 4, the estimated hazard ratio for disease recurrence after therapy in the group of patients with positive SPAIs as compared with the group of patients with negative SPAIs defined by the 11-gene MTTS/PNS signature was 9.259 (95% confidence interval of ratio, 1.545 to 26.07; P = 0.0104). 58% of patients with the positive SPAIs had a disease recurrence within 3 years after therapy, whereas 90% of patients with the negative SPAIs remained relapse-free. Five-year after therapy, 69% of patients with the positive SPAIs had a disease recurrence, whereas 90% of patients with the negative SPAIs remained relapse-free. Based on this analysis, we proposed to identify the group of prostate cancer patients with positive values of the PNS neurosphere-derived 11-gene MTTS/PNS signature as a poor prognosis group and the group of prostate cancer patients with negative values of the 11-gene MTTS/PNS signature as a good prognosis group.

2) Further Analysis of the 11-gene MTTS/PNS Signature

[0108] The identified signature genes were defined based on a strong correlative behavior in multiple independent sets of experimental and clinical samples obtained from two species (mice and human). To test by independent methods the suspected association of the expression of *BMI-1*-pathway target genes with the expression of the *BMI-1* gene product in the context of human cancer cells, we subjected human prostate carcinoma cells to the siRNA-mediated silencing of expression of the endogenous *BMI-1* gene. The PC-3-32 human prostate carcinoma cells were transfected with *BMI-1* or control siRNAs and continuously monitored for mRNA expression levels of *BMI-1* and selected set of genes using RT-PCR and Q-RT-PCR methods (data not shown). RT-PCR and Q-RT-PCR analyses showed that the employed siRNA-mediated *BMI-1*-silencing protocol allowed for ~90% inhibition of the endogenous *BMI-1* mRNA expression. We validated the effect of siRNA-mediated *BMI-1* silencing at the *BMI-1* protein expression level using immunofluorescent analysis. The *BMI-1* silencing was specific since the expression levels of nine un-related transcripts (such as GAPDH, EZH2, and several other genes) were not altered (data not shown). Consistent with the hypothesis that expression of genes comprising the 11-gene MTTS/PNS signature is associated with the expression of the *BMI-1* gene product, mRNA abundance levels of 8 of 11 interrogated *BMI-1*-pathway target genes were altered in the human prostate carcinoma cells with ~90% silenced *BMI-1* gene.

[0109] Reduction of the *BMI-1* mRNA and protein expression in human prostate carcinoma metastasis precursor cells did not alter significantly the viability of adherent cultures grown at the optimal growth condition and in serum starvation experiments (data not shown) and had only modest inhibitory effect on proliferation (~ 25-30% reduction in the number of cells during the 3-day silencing protocol). However, the ability of human prostate carcinoma cells to survive in non-adherent state was severely affected after siRNA-mediated reduction of the *BMI-1* expression. Fluorescence activated cell sorting (FACS) analysis revealed ~ 3-fold increase of apoptosis in the *BMI-1* siRNA-treated human prostate carcinoma cells cultured in non-adherent conditions. These data suggest that human prostate carcinoma cells expressing high level of the *BMI-1* protein are more resistant to apoptosis induced in cells of epithelial origin in response to attachment deprivation (anoikis) and, perhaps, would survive better in blood during metastatic dissemination thus forming a pool of circulatory stress-surviving metastasis precursor cells. Further detail regarding identification of molecular signatures, usage of Pearson coefficients, the Kaplan-Meier survival analysis, and other methods described above is provided in pending U.S. Patent Application Serial Number 10/861,003, filed June 3, 2004, and pending U.S. Patent Application Serial Number 10/660,434, filed September 10, 2003. both of which are hereby incorporated by reference in their entireties.

Expression of the 11-gene MTTS/PNS signature in primary prostate tumors is a predictor of a therapy failure in prostate cancer patients

[0110] To validate a survival prediction model based on the 11-gene MTTS/PNS signature, we tested the prognostic performance of the model in the multiple independent therapy outcome data sets representing five epithelial and five non-epithelial cancers. We divided patients within individual cohorts into a training set, which was used for the cutoff threshold selection and to test the model, and a test set, which was used to evaluate the reproducibility of the classification performance. Using the training set of samples, we selected the prognosis discrimination cut-off value for a signature based on highest level of statistical significance in patient's stratification into poor and good prognosis groups as determined by the log-rank test (lowest P value and highest hazard ratio in the training set). Clinical samples having the Pearson correlation coefficient at or higher than the cut-off value were identified as having the poor prognosis signature. Clinical samples with the Pearson correlation coefficient lower than the cut-off value were identified as having the good prognosis signature. The same discrimination cut off value was then applied to evaluate the reproducibility of the prognostic performance in the test set of patients. Lastly, we applied the model to the entire outcome set using the same cut off threshold to confirm the classification performance. The training and test sets were balanced with respect to the total number of patients, negative and positive therapy outcomes, and the length of survival. We would like to point out that at this stage of the analysis, we did not carry out additional model training, development or optimization steps, except for selecting the prognostic cut off threshold using the training set. We consistently used throughout the study the same MTTS/PNS expression profile as a reference standard to quantify the Pearson correlation coefficients of the individual samples.

[0111] In addition to this analysis, we confirmed the model performance using various sample stratification approaches such as terrain (TRN) clustering, support vector machine (SVM) classification, and weighted survival score algorithm. Finally, we evaluated the therapy outcome predictive power of the 11-gene model in prostate cancer setting using a prognostic test based on an independent method of gene expression analysis, namely quantitative reverse-transcription polymerase chain reaction (Q-RT-PCR) method.

[0112] To further validate the potential clinical utility of the 11-gene MTTS/PNS signature, we evaluated the prognostic power of the 11-gene MTTS/PNS signature applied to an independent set of 79 clinical samples (prostate cancer outcome set 2) obtained from 37 prostate cancer patients who developed recurrence after the therapy and 42 patients who remained disease-free. In this cohort of patients, the Kaplan-Meier survival analysis demonstrated a highly significant difference in the probability that prostate cancer patients would remain disease-free after therapy between the groups with positive and negative SPAIs defined by the 11-gene MTTS/PNS signature. As illustrated in FIG. 5, the estimated hazard ratio for disease recurrence after therapy in the group of patients with positive SPAIs as compared with the group of patients with negative SPAIs defined by the 11-gene MTTS/PNS signature was 3.74 (95% confidence interval of ratio, 3.010 to 25.83; $P < 0.0001$). 67% of patients with the positive SPAIs had a disease recurrence within 3 years after therapy, whereas 70% of patients with the negative SPAIs remained relapse-free. Five-years after therapy, 83% of patients with the positive SPAIs had a disease recurrence, whereas 64% of patients with the negative SPAIs remained relapse-free.

[0113] The standard Kaplan-Meier log-rank statistic assesses the difference in the survival curves, however, it does not account for multiple hypothesis testing and random co-occurrence representing inherent problems of gene expression profiling experiments. In part, we attempted to mitigate this problem by using an alternative biological end-point to the patients' survival during the signature selection process and by applying the survival analysis to a single signature, thus eliminating the multiple comparisons from the survival model building protocol. The MTTS signature is likely to carry many gene expression patterns displaying metastatic phenotype discriminative power that has no relation to the transcriptional program of normal stem cells. One of our main goals was to identify the stem cell signature that is associated with the pluripotency self-renewal phenotype and is embedded into MTTS signature. This approach implies that a candidate marker

signature would have a defined stem cell-like expression profile that can be used in the subsequent follow-up validation analyses as a reference standard to look for expression of a stem cell-like signature in clinical samples.

[0114] To further assess the statistical validity of the 11-gene stem cell-like profile, we performed 1000 random permutations of the 11-gene stem cell profiles randomly selected from the 1973-gene PNS signature. For each random 11-gene stem cell profile we assessed its metastatic phenotype discriminative performance in the TRAMP transgenic mouse model at the discriminative confidence levels of the 11-gene *BMI-1*-pathway MTTS/PNS signature. Only one random 11-gene stem cell profile of the 1000 permutations demonstrated classification power matching the metastatic phenotype discriminative performance of the 11-gene MTTS/PNS signature. We performed 10,000 permutations test to check how likely small 11-gene signatures derived from the large MTTS signature would display high discrimination power to assess the significance at the 0.1% level. We carried out 10,000 permutations of small 11-gene signatures derived from the large 1345-gene MTTS signature and compared their sample stratification power to the 11-gene MTTS/PNS signature. The classification performance cut-off p-values were established by applying two-tailed T-test to the 11-gene MTTS/PNS signature ($p = 0.0005$ for metastasis versus primary prostate cancer data set and $p = 0.026$ for recurrent versus non-recurrent prostate cancer data set). We found that 10,000 permutations generated 7 random 11-gene signatures performing at sample classification level of the 11-gene MTTS/PNS signature.

Cox proportional hazards survival regression analysis

[0115] To ascertain the incremental statistical power of the individual covariates as predictors of therapy outcome and unfavorable prognosis, we performed both univariate and multivariate Cox proportional hazard survival analyses. Several individual gene members of the 11-gene MTTS/PNS signature, such as *MKI67* and *CCNB1*, have been described previously as significant predictors of prognosis and may reflect correlation between proliferative fraction and poor therapy outcome as it has been shown recently for the lymphoma survival predictor signature. However, our analysis appears to indicate that the 11-gene MTTS/PNS signature is a more uniform therapy outcome predictor across the multiple data sets compared to the individual genes (see below) and, perhaps, is a better "integrator" and "sensor" of the biological diversity across the spectrum of human cancers. We performed both univariate and multivariate Cox proportional hazard survival analyses to compare the prognostic performance of the entire stemness signature and individual genes. The results of these analyses are shown in Tables 3 and 4, below.

Table 3. Cox Proportional Hazard Survival Regression Analysis

Prostate Cancer		
Covariates	Statistics	Remarks
GBX2	Chi Square= 1.5817; df=1; p= 0.2085	
MKI67	Chi Square= 9.9016; df=1; p= 0.0017	
CCNB1	Chi Square= 0.1370; df=1; p= 0.7113	
BUB1	Chi Square= 0.9193; df=1; p= 0.3377	
KNTC2	Chi Square= 2.3450; df=1; p= 0.1257	
USP22	Chi Square= 0.1376; df=1; p= 0.7106	
HCFC1	Chi Square= 2.2379; df=1; p= 0.1347	
RNF2	Chi Square= 1.6235; df=1; p= 0.2026	
ANK3	Chi Square= 8.9237; df=1; p= 0.0028	
FGFR2	Chi Square= 7.7985; df=1; p= 0.0052	
CES1	Chi Square= 9.3565; df=1; p= 0.0022	
Signature	Chi Square= 3.9990; df=1; p= 0.0455	
5 Covariates	Chi Square= 26.6628; df=5; p= 0.0001	Signature + 4 genes

6 Covariates	Chi Square= 26.9003; df=6; p= 0.0002	Signature + 5 genes
11 Covariates	Chi Square= 26.9684; df=11; p= 0.0046	11 genes
12 Covariates	Chi Square= 29.2850; df=12; p= 0.0036	Signature + 11 genes
11 Covariates	Chi Square= 50.7039; df=11; p= 0.0000	Signature + 4 genes + 6 clinical
Breast Cancer		
Covariates	Statistics	Remarks
GBX2	Chi Square= 0.0021; df=1; p= 0.9631	
MKI67	Chi Square= 3.7357; df=1; p= 0.0533	
CCNB1	Chi Square= 4.6430; df=1; p= 0.0312	
BUB1	Chi Square= 10.4330; df=1; p= 0.0012	
KNTC2	Chi Square= 15.6837; df=1; p= 0.0001	
USP22	Chi Square= 0.5386; df=1; p= 0.4630	
HCFC1	Chi Square= 0.7418; df=1; p= 0.3891	
RNF2	Chi Square= 0.0360; df=1; p= 0.8495	
ANK3	Chi Square= 2.5573; df=1; p= 0.1098	
FGFR2	Chi Square= 0.2834; df=1; p= 0.5945	
CES1	Chi Square= 0.0477; df=1; p= 0.8272	
Signature	Chi Square= 7.1372; df=1; p= 0.0076	
4 Covariates	Chi Square= 16.4355; df=4; p= 0.0025	Signature + 3 genes
5 Covariates	Chi Square= 16.7995; df=5; p= 0.0049	Signature + 4 genes
11 Covariates	Chi Square= 28.7740; df=11; p= 0.0025	11 genes
12 Covariates	Chi Square= 29.3656; df=12; p= 0.0035	Signature + 11 genes

Table 4. 11 covariates prostate cancer recurrence predictor model

Covariates	Coefficients	Std Errors	Significance, p	Confidence Intervals, Lo95%	Confidence Intervals, Hi95%
Signature	-2.3537	0.9858	0.0170	-4.2858	-0.4215
MKI67	2.2832	0.7823	0.0035	0.7499	3.8166
ANK3	-0.1563	0.7197	0.8280	-1.5670	1.2543
FGFR2	-0.8295	0.4955	0.0941	-1.8007	0.1418
CES1	-1.6403	0.8113	0.0432	-3.2303	-0.0502
PRE RP PSA	0.0493	0.0251	0.0495	0.0001	0.0985
RP GLSN SUM	0.2850	0.2385	0.2322	-0.1825	0.7525
SM	1.0609	0.4648	0.0225	0.1499	1.9720
Sem Ves Inv	0.6016	0.5064	0.2348	-0.3909	1.5941
AGE	0.0311	0.0351	0.3755	-0.0377	0.0999
ECE	0.9296	0.4360	0.0330	0.0751	1.7842

RP, radical prostatectomy; PSA, prostate specific antigen; SM, surgical margins; GLSN SUM, Gleason sum; Sem Ves Inv, seminal vesicle invasion; ECE, extracapsular extension.

[0116] In the univariate analysis prognostic performance of *MKI67* expression as a predictor of therapy outcome varied in different outcome data sets. It was highly significant in the prostate cancer therapy outcome set 2 (MSKCC data set); however, it showed only a trend toward statistical significance in the prostate cancer outcome set 1 (P = 0.1; MIT data set) and breast cancer outcome data set (P = 0.0533). In prostate cancer, the significant prognosis predictors in univariate Cox regression analysis were *MKI67, ANK3, FGFR2, CES1*, and the 11-gene MTTS/PNS signature. In breast cancer, the significant prognosis predictors in univariate analysis were *CCNB1, BUB1, KNTC2*, and the 11-gene MTTS/PNS signature. Thus, our analysis seems to indicate that individual genes demonstrate a variable performance across multiple outcome data sets and we were unable to identify a single gene uniformly predictive of the poor therapy outcome.

[0117] In the multivariate analysis, the most significant prostate cancer recurrence predictor was the model that included 11 covariates [11-gene signature, four individual genes (*MKI67; ANK3; FGFR2; CES1*); and six clinico-pathological features (pre RP Gleason sum; surgical margins; seminal vesicle invasion; age; and extra-capsular extension)]. Interestingly, several covariates such as the 11-gene MTTS/PNS signature, *MKI67, CES1*, pre RP PSA level, surgical margins, and extra capsular extension remained statistically significant prognostic markers in the multivariate analysis. Thus, while prognostic performance of individual gene members of the 11-gene MTTS/PNS signature varied greatly in different outcome data sets, the identified 11-gene MTTS/PNS signature seems to perform as the most consistent predictor of poor therapy outcome across multiple independent outcome data sets comprising over 1,000 clinical samples and representing ten distinct types of human cancer (see below). Yet statistically the best-performing multivariate cancer type-specific model seems to require a combination of calls based on expression levels of individual genes, a gene expression signature, and clinico-pathological covariates.

[0118] We sought to use an alternative statistical metric to further evaluate the prognostic power of the genes comprising the 11-gene MTTS/PNS signature. We implemented the weighted survival score analysis to reflect the incremental statistical power of the individual covariates as predictors of therapy outcome based on a multi-component prognostic model, as illustrated in FIG. 6. Final survival predictor score comprises a sum of scores for individual genes and reflects the relative contribution of each of the eleven genes in the multivariate analysis. The negative weighting values imply that higher expression correlates with longer survival and favorable prognosis, whereas the positive score values indicate that higher expression correlates with poor outcome and shorter survival. Application of the weighted survival predictor model based on a cumulative score of the weighted expression values of eleven genes confirmed the prognostic power of identified 11-gene MTTS/PNS signature in stratification of prostate cancer patients into sub-groups with statistically distinct probability of relapse-free survival after radical prostatectomy.

Expression of the 11-gene MTTS/PNS signature is a predictor of a short relapse-free survival after therapy in prostate cancer patients with an early stage disease

[0119] Identification of patients with high likelihood of poor outcome after therapy would be particularly desirable in a cohort of patients diagnosed with a seemingly localized early stage prostate cancer. Next we determined whether the 11-gene MTTS/PNS signature would be useful in defining sub-groups of patients diagnosed with an early stage prostate cancer and having a statistically significant difference in the likelihood of disease relapse after therapy. In the group of patients diagnosed with the stage 1C or 2A prostate cancer, as shown in FIG. 7, the median relapse-free survival after therapy in the poor prognosis sub-group defined by the 11-gene MTTS/PNS signature was 27 months. In contrast, the median relapse-free survival after therapy in the good prognosis group was 82.4 months. 88 % of patients in the poor prognosis sub-group had a disease recurrence within 5 years after therapy. Conversely, 64 % of patients in the good prognosis sub-group remained relapse-free (FIG. 7). The estimated hazard ratio for disease recurrence after therapy in the poor prognosis sub-group as compared with the good prognosis sub-group of patients defined by the 11-gene MTTS/PNS signature was 3.907 (95% confidence interval of ratio, 2.687 to 34.84; P = 0.0005).

Validation of the prognostic performance of the 11-gene MTTS/PNS signature using a quantitative RT-PCR-based assay

[0120] Routine clinical use of prognostic tests based on microarray-derived gene expression signatures would require the prospective validation study of the utility of identified markers in an experimental setting highly compatible with the state of the art clinical laboratory practice. Since microarray-based assay format is not readily available for application in clinical laboratory, we considered the Q-RT-PCR-based test as an alternative clinically compatible analytical platform suitable for measurements of mRNA expression level of marker genes. Expression of messenger RNA (mRNA) for eleven genes using a set of primers identified in Tables 5 and 6 below and an endogenous control gene (GAPDH) was measured in twenty specimens of primary prostate cancer obtained from patients with documented PSA recurrence within five years after RP and patients who remained disease-free for at least five years after RP (ten patients in each group) by real-time PCR method. As shown in FIG. 8, a prostate cancer therapy outcome test based on measurements of mRNA expression levels of eleven genes using Q-RT-PCR method discriminates prostate cancer patients into subgroups with statistically distinct probability of relapse-free survival after radical prostatectomy.

Table 5. Primer sequences for Q-RT-PCR analysis of the mRNA expression levels of genes comprising the 11-gene MTTS/PNS signature

Gene name	UniGene ID	Sequence (5' - 3')	Amplicon, bp	SEQ ID NO.
GBX2-F	Hs.184945	AAGGCTTCCTGGCCAAAGAG	104	3
GBX2-R		TGACTCGTCTTTCCCTTGCC		4
MKI67-F	Hs.80976	CGCAAACTCTCCTTGTACCATAAT	201	5
MKI67-R		ATAGCGATGTGACATGTGCTTG		6
CCNB1-F	Hs.23960	TGCAGCAGGAGCTTTTTGCT	119	7
CCNB1-R		CCAGGTGCTGCATAACTGGAA		8
BUB1-F	Hs.469649	ACACCATTCCACAAGCTTCCA	123	9
BUB1-R		TGAAGGCACCACCATGTTTTC		10
KNTC2-F	Hs.414407	TGCCAGTGAGCTTGAGTCCTT	136	11
KNTC2-R		TTCAGTCGTGGTTTGCACAAC		12
USP22-F	Hs. 462492	TCAAGTGTGACGATGCCATCA	124	13
USP22-R		CTGACCAGCTGCAGATAAGGCT		14
HCFC1-F	Hs.83634	CCAATGGCATCGAGTCCCT	109	15
HCFC1-R		GTGCCCTTAATGACTCCCACATC		16
RNF2-F	Hs.124186	AGTATTAGCCAGGATCAACAAGCA	104	17
RNF2-R		TCTTGCCTCGCTGCAGTCT		18

ANK3-F	Hs.499725	CCAAGGCTTAGCCTCCATGAA	135	19
ANK3-R		ACTGACCGTTCGCTGTTACGAG		20
FGFR2(1)-F	Hs.533683	CTCCGGCCTCTATGCTTGTACT	114	21
FGFR2(1)-R		CCATCGGTG TCATCCTCATCA		22
FGFR2(2)-F	Hs. 533683	ATAGCAGACTTTGGACTCGCCA	146	23
FGFR2(2)-R		CCGAAGGACCAGACATCACTCT		24
CES1(1)-F	Hs.558865	GGAATTTCCACACTGTCCCCTA	137	25
CES1(1)-R		GGACTTCCACAGGAGTGACATG		26
CES1(2)-F	Hs.558865	TGTTCCTGGACTTGATAGCAGATG	117	27
CES1(2)-R		AGCTTGGACGGTACTGAAACTCA		28

Table 6. Primer sequences for human BMI-1 gene used for Q-RT-PCR analysis[1]

Gene	Orientation	Primer Sequence, 5' - 3'	Product	SEQ ID NO.
Human Bmi-1 outer primers	Sense	ctctgtatttcaatggaagtggaccattcc		29
	Anti-sense	gtatggttcgttacctggagaccagca		30
Human Bmi-1 inner primers	Sense	tcttaagtgcatcacagtcattgctgctg	359 bp	31
	Anti-sense	gatgtccaagttcacaagaccagaccactact		32

[1] Reference: Park, I.-K., Qian, D., Kiel, M., Becker, M.W., Pihalja, M., Weissman

While the Tables above provide examples of primer sequences for Q-RT-PCR analysis of the mRNA expression levels of genes comprising the 11-gene MTTS/PNS signature, one of ordinary skill in the art would recognize that other primer sequences for this PCR analysis of the mRNA expression levels of genes of the 11-gene MTTS/PNS signature are available at a number of sources, such as the National Center for Biotechnology, at http://www.ncbi.nlm.nih.gov/ (e.g., by selecting "UniSTS" from the search window drop down menu for selection of databases to search and by conducting a search for the gene name (e.g., GBX2)) and at Primer3 for the Whitehead Institute for Biomedical Research at http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi.

[0121] The Kaplan-Meier survival analysis demonstrated that application of the 11gene Q-RT-PCR-based prostate cancer therapy outcome test segregates prostate cancer patients into sub-groups with statistically significant difference in the probability to remain relapse-free after the therapy (FIG. 8). The estimated hazard ratio for disease recurrence after therapy in the poor prognosis group of patients as compared with the good prognosis group defined by the test was 21.3 (95%

confidence interval of ratio, 5.741 to 98.39; P < 0.0001). 100% of patients in the poor prognosis group had a disease recurrence within four years after RP, whereas 91% of patients in the good prognosis group remained relapse-free (FIG. 8).

Expression of the 11-gene MTTS/PNS signature predicts metastatic recurrence and poor survival after therapy in breast cancer and lung adenocarcinoma patients diagnosed with an early stage disease

Breast Cancer

[0122] We also sought to investigate whether measurements of expression of the 11-gene MTTS/PNS signature would be informative in the prediction of the patient's prognosis in the group of 97 young women diagnosed with sporadic lymph-node-negative early stage breast cancer (this group comprises of 46 patients who developed distant metastases within 5 years and 51 patients who continued to be disease-free at least 5 years after therapy; they constitute clinically defined poor prognosis and good prognosis groups, correspondingly). Kaplan-Meier analysis indicates that breast cancer patients with tumors displaying a stem cell-like expression profile of the 11-gene MTTS/PNS signature have significantly higher probability to develop distant metastases within 5 years after therapy and therefore can be identified as a poor prognosis sub-group. Median metastasis-free survival after therapy in the poor prognosis sub-group of breast cancer patients defined by the 11-gene MTTS/PNS signature was 26 months. 84 % of patients in the poor prognosis sub-group were diagnosed with distant metastasis within 5 years after therapy. In contrast, 62 % of patients in the good prognosis sub-group remained metastasis-free. As shown in FIG. 9, the estimated hazard ratio for metastasis-free survival after therapy in the poor prognosis sub-group as compared with the good prognosis sub-group of patients defined by the 11-gene MTTS/PNS signature was 3.762 (95% confidence interval of ratio, 3.421 to 20.27; P < 0.0001). Thus, expression pattern of the 11-gene MTTS/PNS signature is strongly predictive of a short post-diagnosis and post-treatment interval to distant metastases in early stage breast cancer patients.

Lung Adenocarcinoma

[0123] Next we asked whether expression analysis of the 11-gene MTTS/PNS signature would be informative in patient's stratification into sub-groups with distinct survival probability after therapy in the group of 125 patients diagnosed with lung adenocarcinoma (34). Similarly to the prostate and breast cancer patients, the Kaplan-Meier analysis shows that patients with tumors displaying a stem cell-like expression profile of the 11-gene MTTS/PNS signature have significantly higher risk of death after therapy and therefore can be defined as a poor prognosis sub-group. Median survival after therapy in the poor prognosis sub-group of lung adenocarcinoma patients defined by the 11-gene MTTS/PNS signature was 15.2 months. In contrast, the median survival after therapy in the good prognosis sub-group was 48.8 months. 100 % of patients in the poor prognosis sub-group died within 3 years after therapy. Conversely, 58 % of patients in the good prognosis sub-group remained alive. As shown in FIG. 10, the estimated hazard ratio for death after therapy in the poor prognosis sub-group as compared with the good prognosis sub-group of patients defined by the 11-gene MTTS/PNS signature was 3.589 (95% confidence interval of ratio, 2.910 to 46.67; P = 0.0005).

[0124] Next we examined whether the 11-gene MTTS/PNS signature would be useful in defining sub-groups of patients diagnosed with an early stage lung adenocarcinoma and having a statistically significant difference in the survival probability after therapy. In the group of patients diagnosed with the stage 1A lung adenocarcinoma, the median survival after therapy in the poor prognosis sub-group defined by the 11-gene MTTS/PNS signature was 49.6 months. 53 % of patients in the poor prognosis sub-group died within 5 years after therapy. In contrast, 92 % of patients remained alive in the good prognosis sub-group. The estimated hazard ratio for death after therapy in the poor prognosis sub-group as compared with the good prognosis sub-group of patients defined by the 11-gene MTTS/PNS signature was 8.909 (95% confidence interval of ratio, 1.418 to 13.12; P = 0.01).

[0125] Based on this analysis we concluded that detection of a stem cell-like expression profile of the 11-gene MTTS/PNS signature in primary tumors from patients diagnosed with the early stage prostate, breast, and lung carcinomas is

associated with a high propensity toward metastatic dissemination and significantly higher risk of poor therapy outcome. Interestingly, therapy outcome in cancer patients diagnosed with other types of epithelial cancers such as ovarian and bladder cancers seems to manifest similar association with distinct patterns of expression of the 11-gene MTTS/PNS signature, as shown in FIGS. 11 and 12.

Expression of the 11-gene MTTS/PNS signature predicts therapy outcome in patients diagnosed with non-epithelial malignancies

[0126] We further sought to analyze whether the 11-gene MTTS/PNS signature would be useful in defining sub-groups of patients diagnosed with non-epithelial cancers and having a statistically significant difference in the survival probability after therapy. Using Kaplan-Meier method, we analyzed the prognostic power of the 11-gene signature in patients diagnosed with diffuse large B-cell lymphoma; mantle cell lymphoma; acute myeloid leukemia; mesothelioma; medulloblastoma; and glioma (see FIG. 13 as one example showing survival of lymphoma patients). Kaplan-Meier analysis demonstrates that a stem cell-like expression profile of the 11-gene MTTS/PNS signature in primary tumors is a consistent powerful predictor of a therapy failure and short survival in cancer patients diagnosed with five distinct types of non-epithelial cancers. Consistent with our findings, an increased *BMI-1* expression in human medulloblastomas was demonstrated in a recent study (van de Vijver, M.J., et al., A gene expression signature as a predictor of survival in breast cancer. N. Engl. J. Med. 2002, 347:1999-2009). Taken together, these data seem to imply the presence of a conserved *BMI-1*-associated pathway(s) similarly engaged in both neural stem cells and a highly malignant subset of human cancers diagnosed in a wide range of organs and uniformly exhibiting a marked propensity toward metastatic dissemination as well as a high probability of unfavorable therapy outcome.

EXAMPLE 2: STEMNESS EXPRESSION SIGNATURES FOR PREDICTING CLINICAL OUTCOME IN PATIENTS

Expression profiles of invasive primary tumors and distant metastatic lesions in a transgenic mouse model of prostate cancer exhibit marked similarity to normal stem cells

[0127] As described above, the emerging concept of cancer stem cells suggests that an engagement of "stemness" genetic pathways in transformed cells may contribute to tumor progression and metastasis of epithelial malignancies. Thus, inappropriate activation of "stemness" genes in cancer cells may be associated with aggressive clinical behavior and increased likelihood of therapy failure. We measured expression levels of ~12,000 genes in primary prostate tumors and distant metastatic lesions at various anatomic sites of six-month old TRAMP mice and defined differentially regulated transcripts by comparison to the gene expression profiles of age-matched wild-type control mice with no evidence of malignant process in the prostate. This analysis identified 276 and 868 genes with increased transcript abundance levels in invasive primary prostate tumors and distant metastatic lesions, respectively.

[0128] To test whether expression profiles of primary and metastatic prostate tumors resemble transcriptional program of stem cells, we compared the genes up-regulated in primary tumors and metastases to the lists of genes enriched in three distinct stem cell types namely neural stem cells, hematopoietic stem cells, and embryonic stem cells (Ivanova, N.B., et al., A stem cell molecular signature. Science 2002, 298:601-604, incorporated herein by reference). Remarkably, the search for union/intersection of lists identified a large number of common genes in each binary comparison, shown in Table 7, below. Most significant similarity was observed for expression profiles of both advanced stage primary prostate tumors and distant metastases and transcripts enriched in neural stem cells. These data are consistent with the hypothesis that tumor progression toward metastatic disease in a transgenic mouse model of prostate cancer occurs to a significant degree within transcriptional space defined by the "stemness" gene expression program.

Table 7. "Stemness" expression profile of transcripts up-regulated in primary and metastatic tumors of the TRAMP transgenic mouse model of prostate cancer.

276 transcripts up-regulated in primary prostate tumors

Stem cell type	Number (%) of common genes
Neural stem cells (NSC)	87 (31.5%)
Embryonal stem cells (ESC)	15 (5.4%)
Hematopoietic stem cells (HSC)	13 (4.7%)
NSC/ESC	88 (31.9%)
NSC/HSC	2 (0.7%)
ESC/HSC	5 (1.8%)
NSC/ESC/HSC	3 (1.1%)
Overall	213 of 276 (77%)

868 transcripts up-regulated in distant metastatic lesions

Stem cell type	Number (%) of common genes
Neural stem cells (NSC)	178 (20.5%)
Embryonal stem cells (ESC)	57 (6.6%)
Hematopoietic stem cells (HSC)	80 (9.2%)
NSC/ESC	192 (22.1%)
NSC/HSC	13 (1.5%)
ESC/HSC	21 (2.4%)
NSC/ESC/HSC	17 (2.0%)
Overall	558 of 868 (64%)

The Table shows that 276 and 868 transcripts up-regulated in primary prostate tumors and distant metastatic lesions, respectively, of six-month old TRAMP mice were compared to genes enriched in neural, embryonic, and hematopoietic stem cells in search for union/intersection of lists.

Altered expression of "stemness" genes in human prostate cancer

[0129] Next we set out to determine whether the phenomenon of resemblance of "stemness" expression profile is relevant to human prostate cancer. We make use of the list of human homologs for murine HSC-related genes defined through the mouse-human homologous pairs search by direct sequence comparison of expressed sequence tags assemblies to identify "stemness" gene sub-sets in multiple clinical and experimental settings pertinent to human prostate cancer. Results of this analysis seem to indicate that the expression of a substantial fraction of genes enriched in stem cells appears altered in various clinical and experimental settings pathophysiologically relevant to human prostate cancer. Overall, 334 of the interrogated 460 human "stemness" genes (73%) were differentially regulated in at least one of the surveyed clinical or experimental settings listed in the Table 8.

Table 8. Number of "stemness" genes differentially regulated in various clinical and experimental settings relevant to human prostate cancer

Type (number) of clinical samples	**Number of "stemness" genes**
Distant prostate cancer metastases (9)	30
Primary prostate tumors (23)	57
Primary prostate tumors (47)	89
Adjacent normal prostate (47)	80
Experimental setting	**Number of "stemness" genes**
Orthotopic xenografts, PC3MLN4	31

Orthotopic xenografts, PC3 & PC3M	46
Prostate cancer cell lines	99
NPEC	77

To identify "stemness" gene sub-sets in multiple clinical and experimental settings pertinent to human prostate cancer, the human "stemness" gene set was compared to genes enriched in metastatic versus primary human prostate tumors, primary prostate tumors versus adjacent normal prostate tissues, and multiple experimental models of human prostate cancer in search for union/intersection of lists for each setting. The human "stemness" gene set was defined from a list of human homologs for murine HSC-related genes defined through the mouse-human homologous pairs search by direct sequence comparison of expressed sequence tags assemblies. In this example, gene expression profiling data derived from the microarray analyses using the Affymetrix U95A GeneChip were utilized in this analysis (460 of the 822 mouse-human homologous pairs).

[0130] Our data appear to indicate that components of a "stemness" transcriptome are frequently altered at the transcript abundance levels in established human prostate cancer cell lines, xenografts, clinical samples of primary prostate tumors as well as distant metastases, suggesting that differences in expression of "stemness" genes may be associated with distinct features of malignant phenotype of human prostate carcinoma cells. To assess the potential clinical relevance of the altered expression of "stemness" genes in prostate tumors, we thought to analyze whether primary prostate tumors with distinct clinical outcome after therapy would exhibit distinct expression profiles of "stemness" genes. We identified four molecular signatures comprising 23, 14, 5, and 16 "stemness" genes (Gene Sets A, B, C, and D, respectively), shown in Tables 9a, 9b, 9c and 9d, that appear to exhibit distinct expression profiles in prostate tumors from patients with recurrent and non-recurrent disease (See FIGS 14, 15, 16, and 17), suggesting that prostate carcinomas with aggressive clinical behavior and adverse outcome after therapy may activate and suppress an opposite spectrum of "stemness" genes compared to the prostate tumors with indolent clinical course of disease and positive therapy outcome.

Table 9a. 23-Gene "Stemness" gene expression signature associated with recurrent prostate cancer (Gene Set A).

Signature 1	**23 genes**		
Gene	Gene Name	GenBank ID	UniGene ID
ENG	Endoglin	X72012	Hs.76753
NRGN	Neurogranin	X99076	Hs.232004
CLECSF2	C-type lectin (activation-induced)	X96719	Hs.85201
EPB41L2	Erythrocyte membrane protein band 4.1-like 2	AF027299	Hs.440387
GART	Phosphoribosylglycinamide synthetase	X54199	Hs.82285
MXD4	MAX dimerization protein 4	AF040963	Hs.511752
PLEKHB2	Pleckstrin homology domain containing	AL120687	Hs.307033 & Hs.512380
RPGR	Retinitis pigmentosa GTPase regulator	U57629	Hs.378949
EST	Homo sapiens cDNA	W28612	Hs.184724
ARHQ	Ras homolog gene family, member Q	AL043108	Hs.442989
MCM5	Minichromosome maintenance deficient 5	X74795	Hs.77171
GORASP2	Golgi reassembly stacking protein 2	AA447263	Hs.6880
SF3A2	Spliceosomal protein SAP-62	L21990	Hs.115232
KIAA0323	KIAA0323	AI494623	Hs.7911

NME2	Non-metastatic cells 2	X58965	Hs.433416
RPL18	Ribosomal protein L18)	L11566	Hs.409634
ACADVL	Very long chain acyl-CoA dehydrogenase	L46590	Hs.437178
IGBP1	Immunoglobulin-binding protein 1	Y08915	Hs.3631
SOX4	SRY-box 4	X70683	Hs.357901
GATA3	GATA-binding protein 3	X58072	Hs.169946
FADS2	Fatty acid desaturase	AL050118	Hs.388164
ITPR1	Type 1 inositol 1,4,5-trisphosphate receptor	D26070	Hs.149900
KLF4	Kruppel-like factor 4	U70663	Hs.376206
Table 9b.	**14-Gene "Stemness" gene expression signature associated with recurrent prostate cancer (Gene Set B).**		
Signature 2	**14 genes**		
Gene	Gene Name	GenBank ID	UniGene ID
ITGA6	Integrin alpha 6B	S66213	Hs.212296
CRHR2	Corticotropin-releasing hormone receptor 2	U34587	Hs.66578
HOXB2	Homeo box B2	X16665	Hs.290432
HOXA10	Homeo box A10	AC004080	Hs.110637
SMARCD2	SWI/SNF complex 60 KDa subunit B (BAF60B)	U66618	Hs.250581
H2AV	Histone H2A.F/Z variant (H2AV)	AW007731	Hs.301005
DKFZP564I052	DKFZP564I052 protein	AL080063	Hs.5364
ITRR1	Inositol 1,4,5-triphosphate receptor, type 1	D26070	Hs.149900
GCS1	Glucosidase I	X87237	Hs.83919
TGOLN2	Trans -golgi network protein 2	AF027516	Hs.14894
APS	Adaptor protein with pleckstrin homology and src homology 2	AB000520	Hs.371366
GLA	Galactosidase, alpha	U78027	Hs.69089
EST	Protein with strong similarity to A48043	H10776	Hs.107374
MAFF	V-maff musculoaponeurotic fibrosarcoma oncogene homolog F	AL021977	Hs.460889
Table 9c.	**5-Gene "Stemness" gene expression signature associated with recurrent prostate cancer (Gene Set C).**		
Signature 3	**5 genes**		
Gene	Gene Name	GenBank ID	UniGene ID
NRGN	Neurogranin	X99076	Hs.232004
RGS3	Regulator of G-protein □ignaling 3	U27655	Hs.82294
EDIL3	EGF-like repeats and discoidin I-like domains	U70312	Hs.441044
GPR56	G protein-coupled receptor 56	AJ011001	Hs.6527
ITRR1	Inositol 1,4,5-triphosphate receptor, type 1	D26070	Hs.149900

Table 9d.	16-Gene "Stemness" gene expression signature associated with recurrent prostate cancer (Gene Set D).		
Signature 4	16 genes		
Gene	Gene Name	GenBank ID	UniGene ID
LYRIC	LYRIC/3D3	AA398463	Hs.377155
TMSB10	Thymosin, beta 10	M92383	Hs.446574
ZNF183	Zinc finger protein 183	X98253	Hs.64794
PRKCBP1	Protein kinase C-binding protein 1	W22296	Hs.37372 & Hs.191990
ALG3	Asparagine-linked glycosylation 3 homolog	Y09022	Hs.153591
B4GALT4	Beta-1,4-galactosyltransferase	AF038662	Hs.13225
ERCC1	Excision repair cross-complementing 1	M13194	Hs.435981
PTPRK	Protein tyrosine phosphatase, receptor type	L77886	Hs.354262
POU2F2	POU domain, class 2, transcriprion factor 2	M36542	Hs.1101
NFKBIA	NFKB gene enhancer in B-cells inhibitor, alpha	M69043	Hs.81328
Unknown	Homo sapiens cDNA	N48190	Hs.22243
GEM	GTP-binding protein	U10550	Hs.79022
PDE4B	Phosphodiesterase 4B	L20971	Hs.188
RBPMS	RNA-binding protein with multiple splicing	D84110	Hs.195825
GSRP1	Cysteine and glycine-rich protein 1	M33146	Hs.108080
MEIS1	Myeloid ecotropic viral integration site 1 homolog	U85707	Hs.170177

Affymetrix probe ID numbers for the probes corresponding to each of the genes shown in Tables 9a, 9b, 9c, and 9d, and from the Affymetrix probe set U95Av2 can be found at http://www.affymetrix.com/products/arrays/specific/hgu95.affx on the GENECHIP® Human Genome U95 set using the "Array Finder" and either the GenBank ID or Unigene ID as an identifier with which to conduct the search. In addition, a table showing all probes in the U95 probe set (including each probe ID and the corresponding gene, and other details) can be found at https://www.affymetrix.com/analysis/netaffx/showresults.affx.

Prognostic value of "stemness" gene expression signatures

[0131] To further examine the potential clinical utility of the altered expression of "stemness" genes in human prostate cancer, we examined whether the assessment of expression profiles of "stemness" signatures in individual prostate tumors would assist in stratification of prostate cancer patients at the time of diagnosis into sub-groups with statistically distinct likelihood of disease recurrence after radical prostatectomy. We evaluated the prognostic power of each identified "stemness" signature based on ability to segregate the patients with recurrent and non-recurrent prostate tumors into distinct sub-groups. To assess a potential prognostic relevance of individual "stemness" signatures, we calculated a Pearson correlation coefficient for each of 21 tumor samples of the outcome set 1 by comparing the expression profiles of individual samples to the "average" expression profile of recurrent versus non-recurrent tumors (14-gene signature or gene set B) or "stemness" expression profiles of relevant experimental or clinical samples (FIGS. 14, 15, 16, 17 and Table 9b). Based on expected correlation of expression profiles of identified "stemness" signatures with recurrent clinical behavior of prostate

cancer, we named the corresponding correlation coefficients calculated for individual samples the "stemness" phenotype association indices (SPAIs).

[0132] To evaluate the prognostic power of identified "stemness" gene expression signatures, we performed the Kaplan-Meier survival analysis using as a clinical end-point disease-free interval (DFI) after therapy in prostate cancer patients with positive and negative SPAIs. The Kaplan-Meier survival curves showed a highly significant difference in the probability that prostate cancer patients would remain disease-free after therapy between the groups with positive and negative SPAIs defined by the "stemness" signatures (FIGS 18, 19, 20, and 21), suggesting that patients with positive SPAIs exhibit a poor outcome signature whereas patients with negative SPAIs manifest a good outcome signature. The estimated hazard ratio for disease recurrence after therapy in the group of patients with positive SPAIs as compared with the group of patients with negative SPAIs defined by the 23-gene "stemness" signature or gene set A (Table 9a, and FIG. 18) was 30.06 (95% confidence interval of ratio, 20.14 to 800.4; $P < 0.0001$). 100% of patients with the positive SPAIs had a disease recurrence within 3 years after therapy, whereas 100% of patients with the negative SPAIs remained relapse-free at least 3 years (FIG. 18). Five-year after therapy, 100% of patients with the positive SPAIs had a disease recurrence, whereas 92% of patients with the negative SPAIs remained relapse-free (FIG. 18). Based on this analysis, we propose to identify the group of prostate cancer patients with positive "stemness" signatures as a poor prognosis group and the group of prostate cancer patients with negative "stemness" signatures as a good prognosis group.

[0133] Theoretically, the recurrence predictor algorithm based on a combination of signatures should be more robust than a single predictor signature, particularly during the validation analysis using an independent test cohort of patients. We therefore analyzed whether a combination of the four "stemness" signatures would perform in the patient's classification test with similar accuracy as the individual signatures. The Kaplan-Meier survival analysis (FIG. 22) showed that the median relapse-free survival after therapy of patients in the poor prognosis group (defined as having two or more positive "stemness" signatures) was 26 months. 89% of patients in the poor prognosis group had a disease recurrence within 5 years after therapy, whereas 100% of patients in the good prognosis group (defined as having 3 or 4 negative "stemness" signatures) remained relapse-free (FIG. 22; $P < 0.0001$). Using "stemness" signature algorithm, all eight patients who developed disease recurrence after therapy were correctly classified into poor prognosis group.

[0134] To further validate the potential clinical utility of identified "stemnesss" signatures, we evaluated the prognostic power of signatures applied to an independent set of 79 clinical samples (outcome set 2) obtained from 37 prostate cancer patients who developed recurrence after the therapy and 42 patients who remained disease-free. The Kaplan-Meier survival analysis demonstrated that all four individual "stemness" signatures segregate prostate cancer patients into poor and good prognosis sub-groups with statistically significant difference in the probability to remain relapse-free after the therapy.

[0135] Next we determined whether a combination of the four "stemness" signatures would perform in the patient's classification test with similar accuracy as the individual signatures. The Kaplan-Meier survival analysis showed that the median relapse-free survival after therapy of patients in the poor prognosis group (defined as having four positive "stemness" signature) was 6 months (see FIGS. 23 and 24). 80% of patients in the poor prognosis group had a disease recurrence within one year after therapy, whereas 92% of patients in the good prognosis group (defined as having 3 or 4 negative "stemness" signatures) remained relapse-free. All patients in the poor prognosis group had a disease recurrence within 3 years after therapy, whereas 80% of patients in the good prognosis group remained relapse-free at least 3 years. The estimated hazard ration for disease recurrence after therapy in the poor prognosis group of patients as compared with the good prognosis group of patients defined by the recurrence predictor algorithm was 9.172 (95% confidence interval of ratio, 47.79 to 5484; $P < 0.0001$).

[0136] The Kaplan-Meier survival analysis identified in this cohort of patients a group with an intermediate prognosis. The median relapse-free survival after therapy of patients in the intermediate prognosis group defined by the "stemness" algorithm as having 2 or 3 positive signatures was 49.4 months (see FIGS. 23 and 24). 58 % of patients in the intermediate

prognosis group had a disease recurrence within 3 years after therapy, whereas 80 % of patients in the good prognosis group remained relapse-free. 45 % of patients in the intermediate prognosis group had a disease recurrence within 5 years after therapy, whereas 78 % of patients in the good prognosis group remained relapse-free. The estimated hazard ration for disease recurrence after therapy in the poor prognosis group as compared with the good prognosis group of patients defined by the recurrence predictor algorithm was 2.832 (95% confidence interval of ratio, 1.475 to 6.281; P = 0.0026). Overall, the application of the "stemness" recurrence predictor algorithm allowed accurate stratification into poor and intermediate prognosis groups 82 % of patients who failed the therapy within one year after prostatectomy.

[0137] To further ascertain the potential significance of an aberrant expression of "stemness' genes in human prostate cancer, we analyzed the frequency of actual disease recurrence in prostate cancer patients with distinct "stemness" gene expression profiles. This analysis clearly showed that the sub-group of patients with four and three positive "stemness" signatures had highly aggressive malignant disease even at the early stage of progression: 100% of stage 1C patients in this sub-group were diagnosed with disease recurrence after radical prostatectomy. Overall, 76% of patients in this sub-group had recurrent disease and 48% of patients were diagnosed with recurrence within one year after prostatectomy. In contrast, 79 % of patients with four negative "stemness" signatures remained disease-free and only 5% had recurrence within one year after surgery.

[0138] In summary, our analysis seems to indicate that expression of genes identified as components of "stemness" transcriptome is frequently altered in prostate cancer, suggesting that prostate cancer progression occurs at least in part within transcriptional space activated in normal stem cells. One of the hallmark biological features of normal stem cells is the ability to fuse spontaneously in vitro and in vivo with other cell types leading to formation of reprogrammed viable somatic cell hybrids (Vassilopoulos, G., Wang, P.-R., Russell, D.W. Transplanted bone marrow regenerates liver by cell fusion. Nature 2003, 422:901-904; Alvarez-Dolado, M., et al., Fusion of bone-marrow-derived cells with Purkinje neurons, cardiomyocytes and hepatocytes. Nature 2003, 425:968-973; Weimann, J.M., et al., Stable reprogrammed heterokaryons form spontaneously in Purkinje neurons after bone marrow transplant. Nature Cell biology 2003, 5:959-966; LaTulippe, E., et al., Comprehensive gene expression analysis of prostate cancer reveals distinct transcriptional programs associated with metastasis. Cancer Res. 2002, 62:4499-4506, incorporated herein by reference). It would be of interest to study how cancer cells co-opt "stemness" transcriptome into progression pathways and whether some human carcinomas could attract stem cells by mimicking a stem cell "niche" microenvironment thus directly engaging normal stem cells into malignant process.

[0139] While the invention has been particularly shown and described with reference to a preferred embodiment and various alternate embodiments, it will be understood by persons skilled in the relevant art that various changes in form and details can be made therein without departing from the spirit and scope of the invention. All references, issued patents and patent applications cited within the body of the instant specification are hereby incorporated by reference in their entirety, for all purposes.

REFERENCES CITED

1 Al-Hajj, M., Wicha, M.S., Benito-Hernandez, A., Morrison, S.J., Clarke, M.F. 2003. Prospective identification of tumorigenic breast cancer cells. Proc. Natl. Acad. Sci. USA 100:3983-3988.

2 Alkema, M.J., Jacobs, H., van Lohuizen, M., Berns, A. 1997. Perturbation of B and T cell development and predisposition to lymphomagenesis in Eμ-Bmi-1 transgenic mice require the Bmi-1 RING finger. Oncogene 15:899-910.

3 Alvarez-Dolado, M., Pardal, R., Garcia-Verdugo, J.M., Fike, J.R., Lee, H.O., Pfeffer, K., Lois, C., Morrison, S.J., Alvarez-Buylla, A. 2003. Fusion of bone-marrow-derived cells with Purkinje neurons, cardiomyocytes and hepatocytes. Nature 425:968-973.

4 Ausubel, F.M. et al., Current Protocols in Molecular Biology, Volume 2, 1997, pp. 11.12.1-11.12.9.

5 Ausubel, F.M. et al., Current Protocols in Molecular Biology, Volume 2, 1997, pp. 11.4.1-11.11.5.

6 Baron, V., De Gregorio, G., Krones-Herzig, A., Virolle, T., Calogero, A., Urcis, R., Mercola, D. 2003. Inhibition of Egr-1 expression reverses transformation of prostate cancer cells in vitro and in vivo. Oncogene 22:4194-4204.

7 Chang, H.Y., Sneddon, J.B., Alizadeh, A.A., Sood, R., West, R.B., et al. (2004). Gene expression signature of fibroblast serum response predicts human cancer progression: Similarities between tumors and wounds. PLOS Biology 2:1-9.

8 Dick, J.E. 2003. Self-renewal writ in blood. Nature 423:231-233.

9 Dimri, G.P., Martinez, J.-L., Jacobs, J.J.L., Keblusek, P., Itahana, K., van Lohuizen, M., Campisi, J., Wazer, D.E., Band, V. 2002. The Bmi-1 oncogene induces telomerase activity and immortalizes human mammary epithelial cells. Cancer Res. 62:4736-4745.

10 Gingrich, J.R., Barrios, R.J., Morton, R.A., Boyce, B.F., DeMayo, F.J., Finegold, M.J., Agelopoulou, R., Rosen, J.M., Greenberg, N.M. 1996. Metastatic prostate cancer in a transgenic mouse. Cancer Res. 56:4096-4102.

11 Glinsky, G.V., Krones-Herzig, A., Glinskii, A.B., Gebauer, G. 2003. Microarray analysis of xenograft-derived cancer cell lines representing multiple experimental models of human prostate cancer. Molecular Carcinogenesis 37:209-221.

12 Glinsky, Gennadi V. et al, Microarray analysis identifies a death-from-cancer signature predicting therapy failure in patients with multiple types of cancer, J. Clin. Invest. 2005, 1:115(6):1503-1521

13 Haupt, Y., Bath, M.I., Harris, A.W., Adams, J.M. 1993. BMI-1 transgene induces lymphomas and collaborates with Myc in tumorigenesis. Oncogene 8:3161-3164.

14 Ivanova, N.B., Dimos, J.T., Schaniel, C., Hackney, J.A., Moore, K.A., Lemischka, I.R. 2002. A stem cell molecular signature. Science 298:601-604.

15 Lamb, J., Ramaswamy, S., Ford, H.L., Contreras, B., Martinez, R.V., et al. 2003. A mechanism of cyclin D1 action encoded in the patterns of gene expression in human cancer. Cell 114:323-334.

16 LaTulippe, E., Satagopan, J., Smith, A., Scher, H., Scardino, P., Reuter, V., Gerald, W.L. 2002. Comprehensive gene expression analysis of prostate cancer reveals distinct transcriptional programs associated with metastasis. Cancer Res. 62:4499-4506.

17 Lessard, J. and Sauvageau, G. 2003. BMI-1 determines the proliferative capacity of normal and leukaemic stem cells. Nature 423:255-260.

18 Lessard, J., Baban, S., Sauvageau, G. 1998. Stage-specific expression of polycomb group genes in human bone marrow cells. Blood 91:1216-1224.

19 Magee, J.A., Araki, T., Patil, S., Ehrig, T., True, L., Humphrey, P.A., Catalona, W.J., Watson, M.A., Milbrandt, J. 2001. Expression profiling reveals hepsin overexpression in prostate cancer. Cancer Res. 61:5692-5696.

20 Molofsky, A.V., Pardal, R., Iwashita, T., Park, I.-K., Clarke, M.F., Morrison, S.J. 2003. Bmi-1 dependence distinguishes neural stem cell self-renewal from progenitor proliferation. Nature 425:962-967.

21 Pardal, R., Clarke, M.F., Morrison, S.J. 2003. Applying the principle of stem-cell biology to cancer. Nature Review Cancer 3:895-902.

22 Park, I.-K., Qian, D., Kiel, M., Becker, M.W., Pihalja, M., Weissman, I.L., Morrison, S.J., Clarke, M.F. Bmi-1 is required for maintenance of adult self-renewing haematopoietic stem cells. 2003. Nature 423:302-305.

23 Raaphorst, F.M. Vermeer, M., Fieret, E., Blokzijl, T., Mommers, E., Buerger, H., Packeisen, J., Sewalt, R.A., Otte, A.P., van Diset, P.J. 2003. Poorly differentiated breast carcinoma is associated with increased expression of the human polycomb group EZH2 gene. Neoplasia 5:481-488.

24 Smalley, M. and Ashworth, A. Stem cells and breast cancer: a field in transit. 2003. Nature Review Cancer 3:832-844.

25 van de Vijver, M.J., He, Y.D., van 't Veer, L.J., et al. 2002. A gene expression signature as a predictor of survival in breast cancer. N. Engl. J. Med. 347:19992009.

26 Vassilopoulos, G., Wang, P.-R., Russell, D.W. 2003. Transplanted bone marrow regenerates liver by cell fusion. Nature 422:901-904.

27 Vonlanthen, S., et al. 2001. The Bmi-1 oncoprotein is differentially expressed in non-small-cell lung cancer and correlates with INK4A-ARF locus expression. Br. J. Cancer 84:1372-1376.

28 Weimann, J.M., Johansson, C.B., Trejo, A., Blau, H.M. 2003. Stable reprogrammed heterokaryons form spontaneously in Purkinje neurons after bone marrow transplant. Nature Cell Biology 5:959-966.

CLAIMS

What is claimed is:

1. A kit for determining expression of at least three genes selected from the group consisting of *GBX2*, *MKI67*, *CCNB1*, *BUB1*, *KNTC2*, *USP22*, *HCFC1*, *RNF2*, *ANK3*, *FGFR2*, and *CES1*, and mouse homologs thereof, comprising:
a set of probes to specifically detect expression of said at least three genes and that specifically do not detect expression of other genes, and wherein said set of probes comprise nucleic acids or antibodies.

2. The kit of claim 1, wherein said set of probes are nucleic acids capable of hybridizing under normal stringency conditions to RNA species transcribed from said at least three genes or to cDNA species derived from said RNA species.

3. The kit of claim 2, wherein said set of probes are PCR primers.

4. The kit of claim 3, wherein said PCR primers are at least three pair of primers selected from the group consisting of SEQ. ID NO: 3, SEQ. ID NO: 4, SEQ. ID NO: 5, SEQ. ID NO: 6, SEQ. ID NO: 7, SEQ. ID NO: 8, SEQ. ID NO: 9, SEQ. ID NO: 10, SEQ. ID NO: 11, SEQ. ID NO: 12, SEQ. ID NO: 13, SEQ. ID NO: 14, SEQ. ID NO: 15, SEQ. ID NO: 16, SEQ. ID NO: 17, SEQ. ID NO: 18, SEQ. ID NO: 19, SEQ. ID NO: 20, SEQ. ID NO: 213, SEQ. ID NO: 22, SEQ. ID NO: 23, SEQ. ID NO: 24, SEQ. ID NO: 25, SEQ. ID NO: 26, SEQ. ID NO: 27, and SEQ. ID NO: 28.

5. The kit of claim 2, wherein said kit comprises a solid phase.

6. The kit of claim 5, wherein said set of probes consists of at least three probe sets selected from the group consisting of Affymetrix HG-U95Av2 probe set 33688_at, Affymetrix HG-U95Av2 probe set 418_at, Affymetrix HG-U95Av2 probe set 34736_at, Affymetrix HG-U95Av2 probe set 41081_at, Affymetrix HG-U95Av2 probe set 40041_at, Affymetrix HG-U95Av2 probe set 39866_at, Affymetrix HG-U95Av2 probe set 37910_at, Affymetrix HG-U95Av2 probe set 33484_at, Affymetrix HG-U95Av2 probe set 36967_g_at, Affymetrix HG-U95Av2 probe set 1143_s_at, Affymetrix HG-U95Av2 probe set 37203_at, Affymetrix HG-U133A probe set 210560_at, Affymetrix HG-U133A probe set 212022_s_at, Affymetrix HG-U133A probe set 214710_s_at, Affymetrix HG-U133A probe set 216277_at, Affymetrix HG-U133A probe set 204162_at, Affymetrix HG-U133A probe set 216964_at, Affymetrix HG-U133A probe set 202473_x_at, Affymetrix HG-U133A probe set 205215_at, Affymetrix HG-U133A probe set 209442_x_at, Affymetrix HG-U133A probe set 208228_s_at, Affymetrix HG-U133A probe set 209616_s_at, Affymetrix MG-U74A probe set 94200_at, Affymetrix MG-U74A probe set 99457_at, Affymetrix MG-U74A probe set 160159_at, Affymetrix MG-U74A probe set 104097_at, Affymetrix MG-U74A probe set 93441_at, Affymetrix MG-U74A probe set 97960_at, Affymetrix MG-U74A probe set 100901_at, Affymetrix MG-U74A probe set 93164_at, Affymetrix MG-U74A probe set 98477_s_at, Affymetrix MG-U74A probe set 93090_at, and Affymetrix MG-U74A probe set 101538_i_at.

7. The kit of claim 1, wherein said at least three genes are *CCNB1*, *BUB1*, *KNTC2*, or the mouse homologs thereof.

8. The kit of claim 1, wherein said kit is a kit for determining expression of *MKI67*, *ANK3*, *FGFR2*, and *CES1*, or the mouse homologs thereof, and said set of probes specifically detects expression of *MKI67*, *ANK3*, *FGFR2*, and *CES1*, or the mouse homologs thereof.

9. The kit of claim 1, wherein said kit is a kit for determining expression of *GBX2*, *MKI67*, *CCNB1*, *BUB1*, *KNTC2*, *USP22*, *HCFC1*, *RNF2*, *ANK3*, *FGFR2*, and *CES1*, or the mouse homologs thereof, and said set of probes specifically detects expression of *GBX2*, *MKI67*, *CCNB1*, *BUB1*, *KNTC2*, *USP22*, *HCFC1*, *RNF2*, *ANK3*, *FGFR2*, and *CES1*, or the mouse homologs thereof.

10. A method for predicting a clinical outcome for a disease state in a subject, comprising:
obtaining a sample from said subject;
determining from said sample a set of gene expression measurements for at least three genes selected from the group

consisting of *GBX2, MKI67, CCNB1, BUB1, KNTC2, USP22, HCFC1, RNF2, ANK3, FGFR2,* and *CES1*, or the mouse homologs thereof; and

determining a correlation coefficient between said set of gene expression measurements and a reference standard set of gene expression measurements obtained by comparing expression values from a stem cell and from a tumor cell for said set of genes, wherein the sign of said correlation coefficient is predictive of said clinical outcome for said disease state.

11. The method of claim 10, wherein said stem cell is a peripheral nervous system neurosphere.

12. The method of claim 10, wherein said tumor cell is a metastatic prostate tumor cell.

13. The method of claim 10, wherein said disease state is cancer.

14. The method of claim 13, wherein said cancer is selected from the group consisting of prostate cancer, breast cancer, lung cancer, ovarian cancer, bladder cancer, lymphoma, mantle cell lymphoma, mesothelioma, medulloblastoma, glioma, and acute myeloid leukemia.

15. The method of claim 13, wherein said at least three genes are *CCNB1, BUB1, KNTC2,* or the mouse homologs thereof.

16. The method of claim 13, wherein said set of gene expression measurements are expression measurements of *MKI67, ANK3, FGFR2,* and *CES1*, or the mouse homologs thereof.

17. The method of claim 13, wherein said set of gene expression measurements are expression measurements of *GBX2, MKI67, CCNB1, BUB1, KNTC2, USP22, HCFC1, RNF2, ANK3, FGFR2,* and *CES1*, or the mouse homologs thereof.

18. The method of claim 13, wherein said clinical outcome is selected from the group consisting of recurrence, therapy failure, likelihood of metastasis, likelihood of distant metastasis, disease free survival, invasiveness, and likelihood of survival at a predetermined time period.

19. The method of 14, wherein said cancer is prostate cancer.

20. The method of claim 19, further comprising analyzing a clinico-pathological feature selected from the group consisting of a pre-radical prostatectomy Gleason sum, a surgical margin evaluation, a seminal vesicle invasion, an age, and an extra-capsular extension.

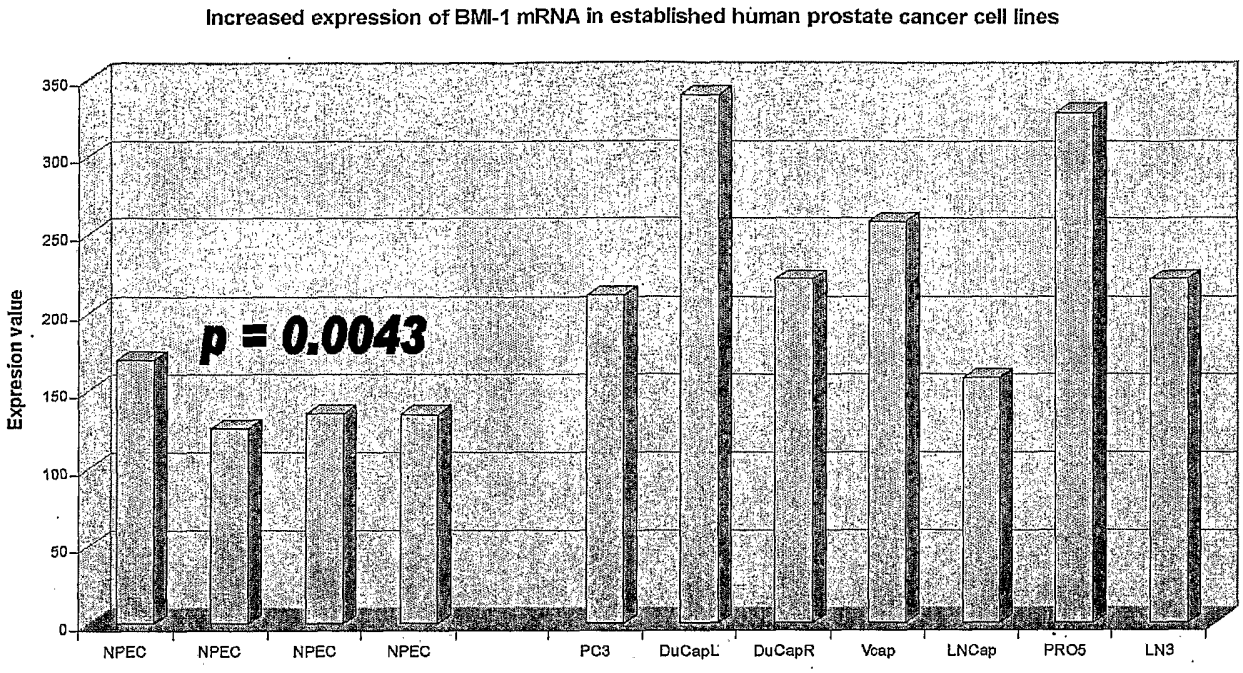


FIGURE 1

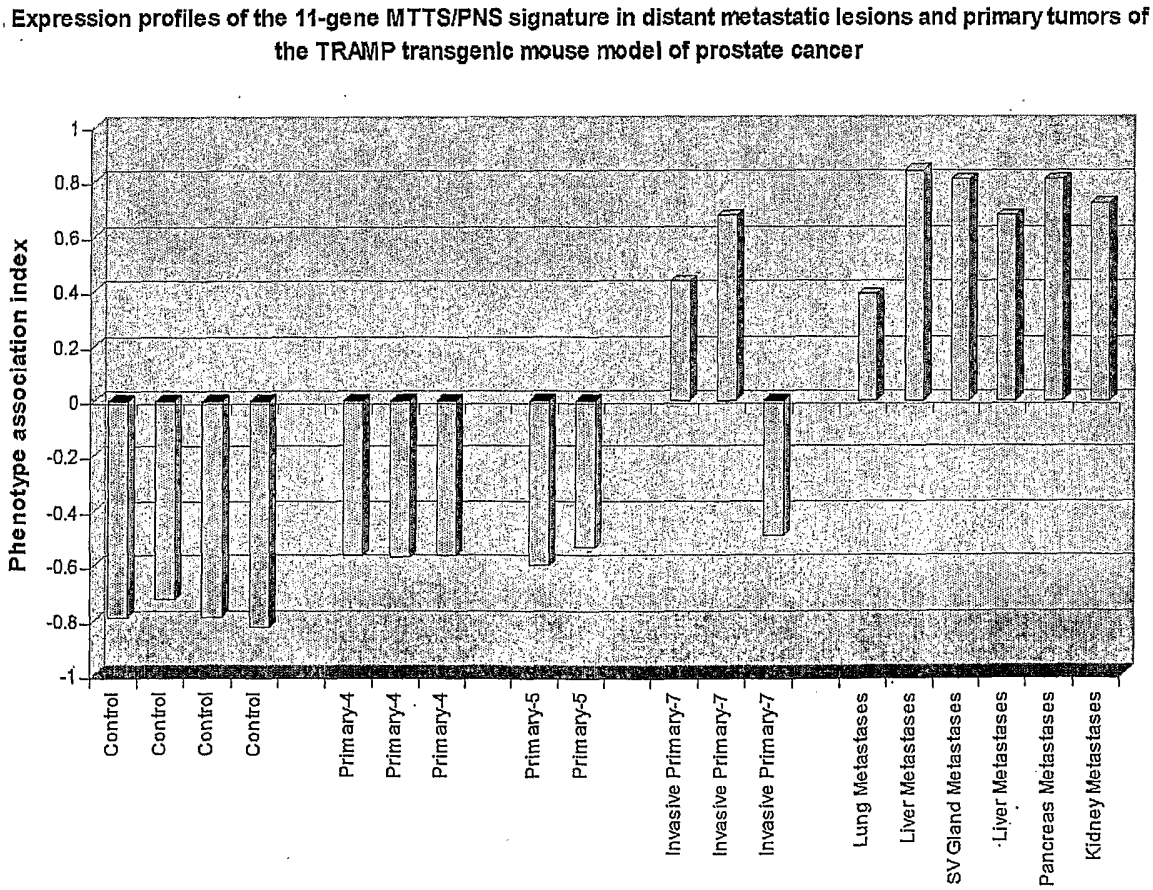


FIGURE 2

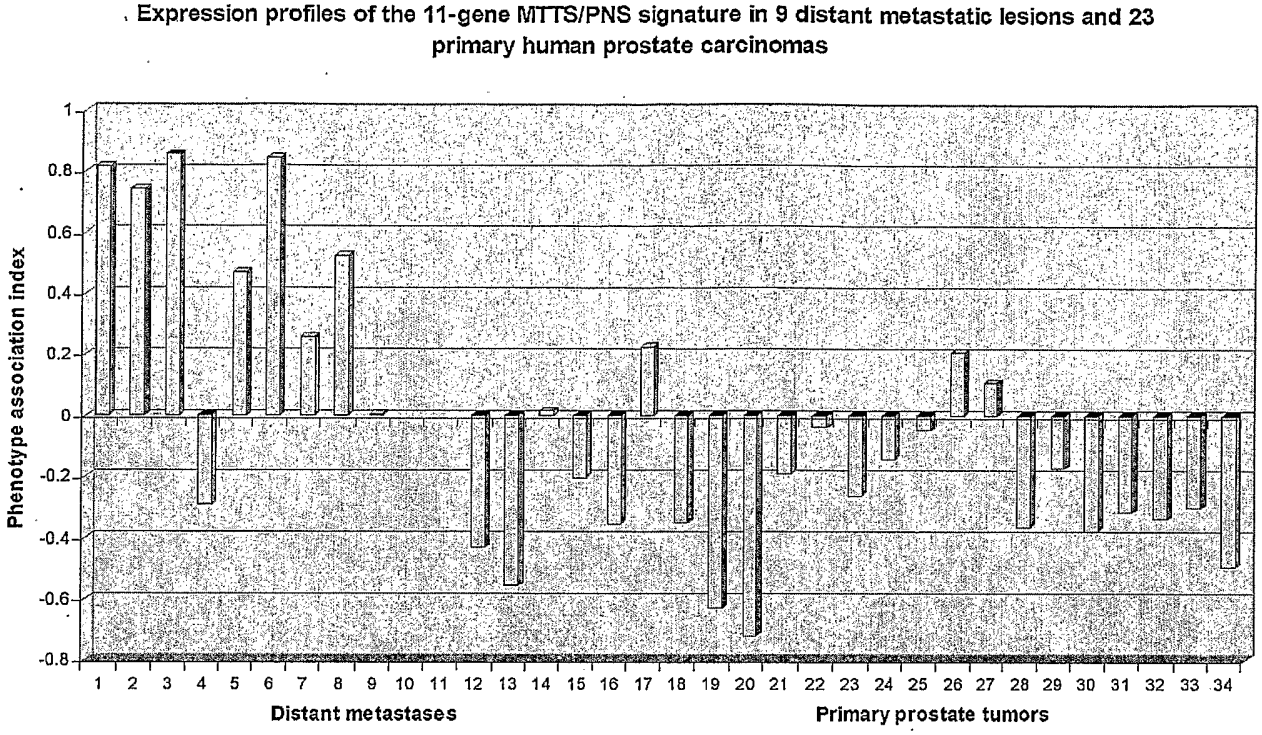


FIGURE 3

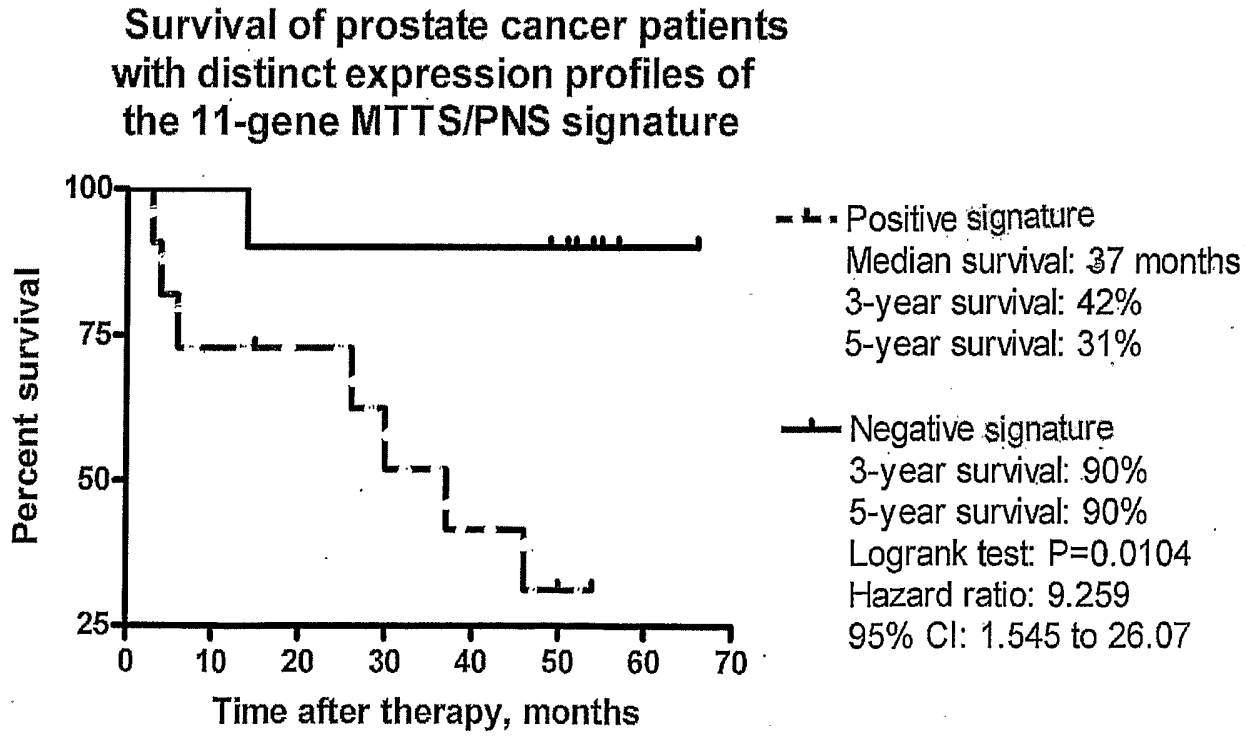


FIGURE 4

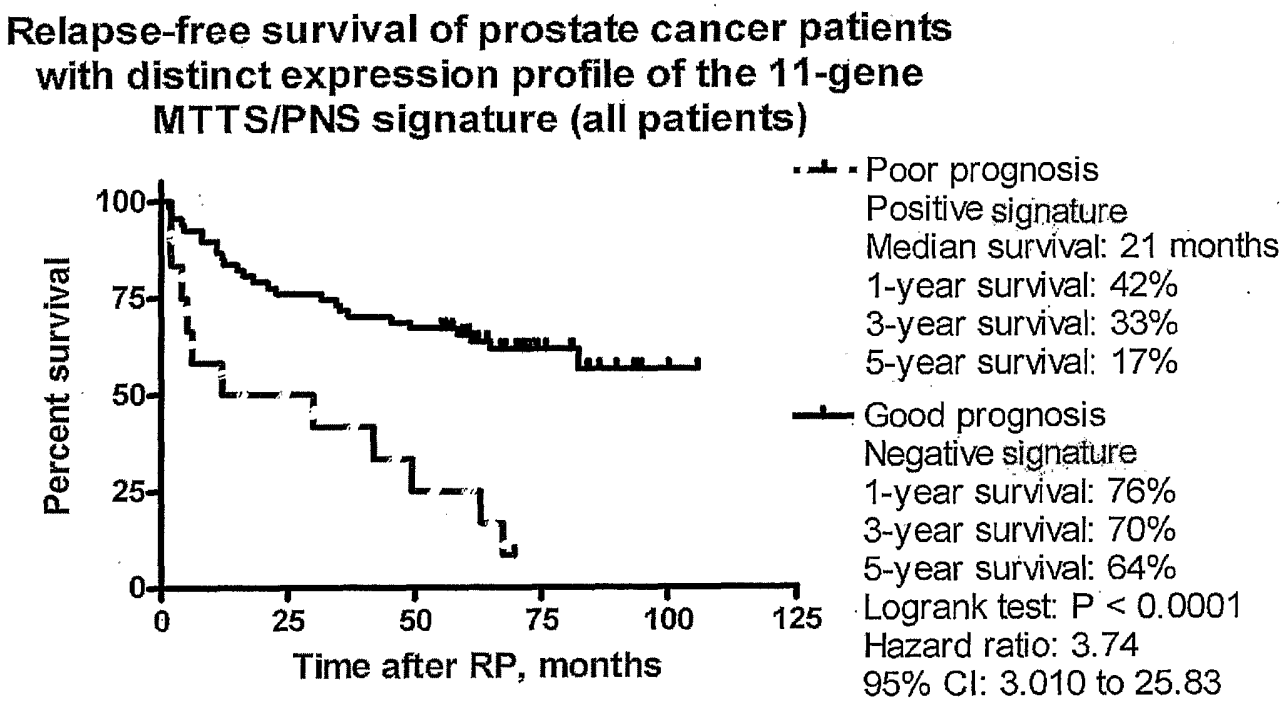


FIGURE 5

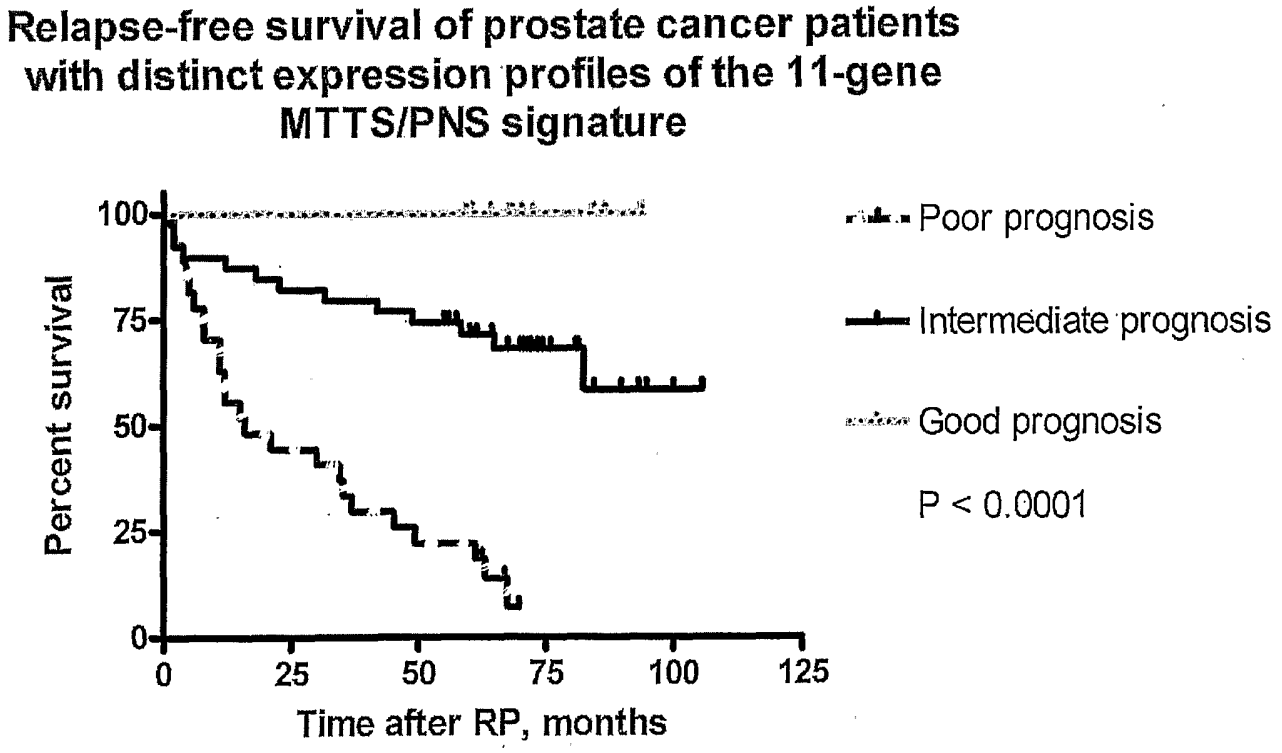


FIGURE 6

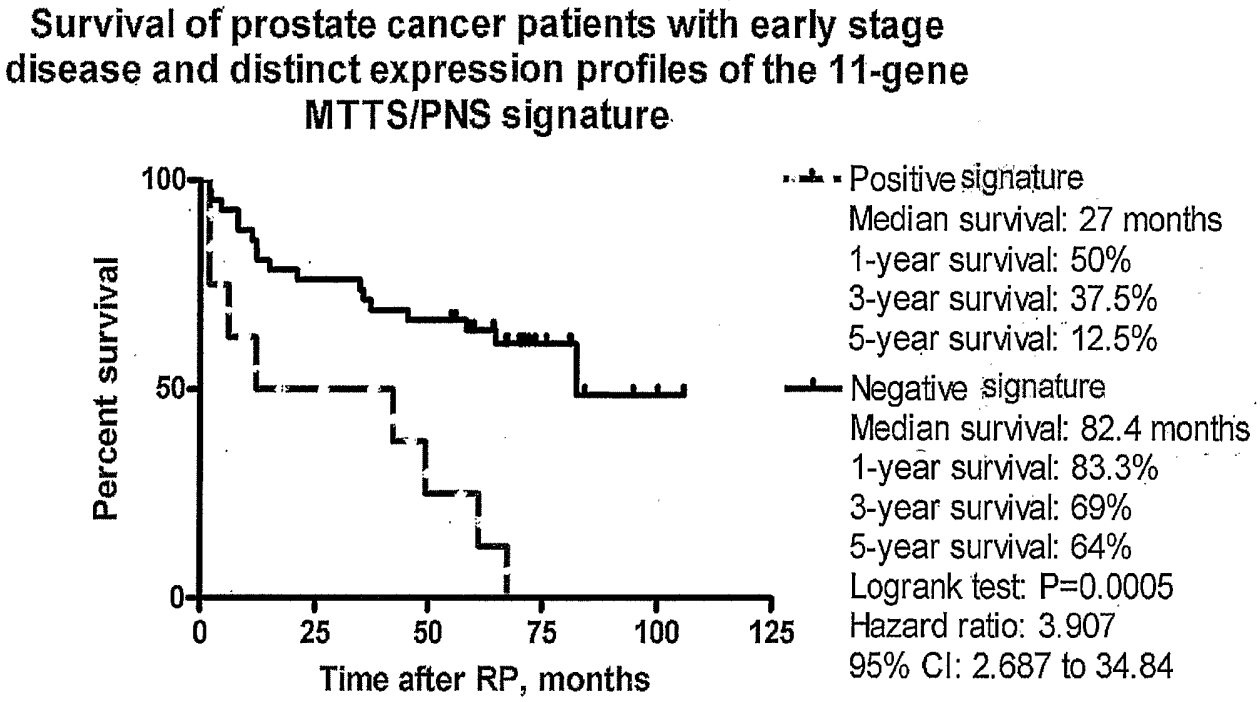


FIGURE 7

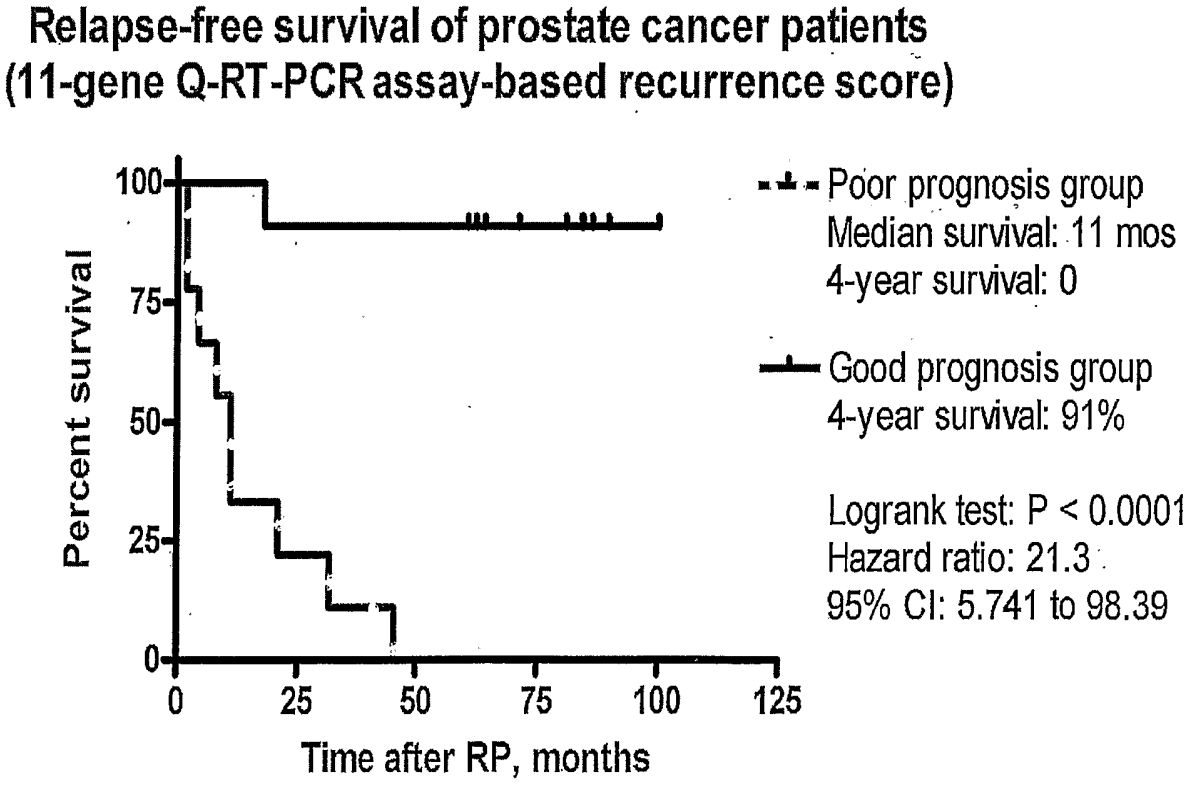


FIGURE 8

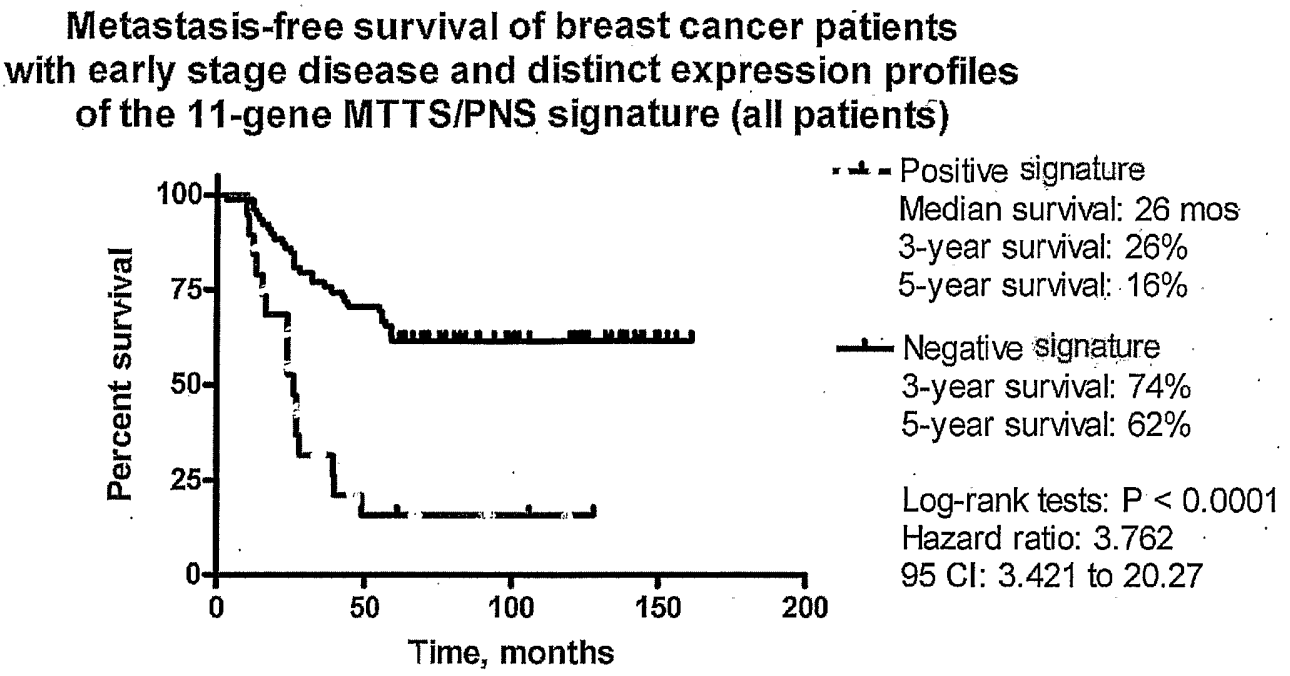


FIGURE 9

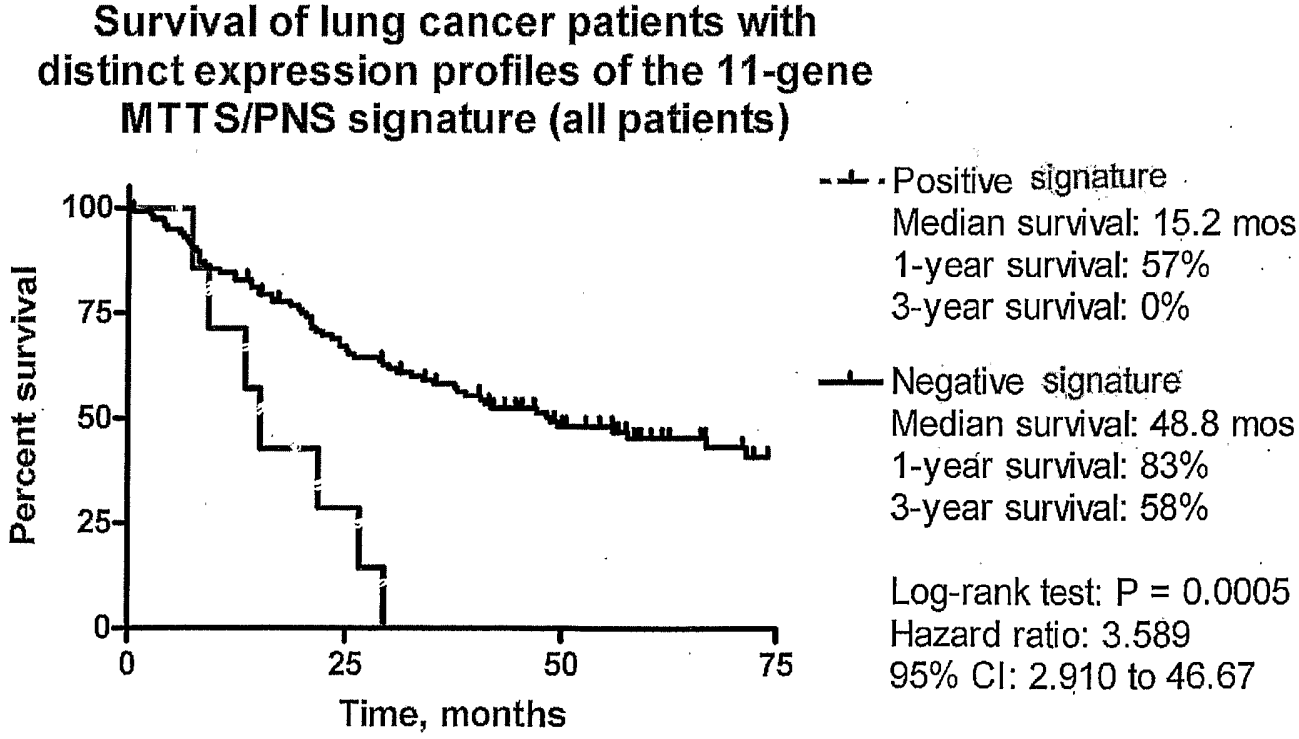


FIGURE 10

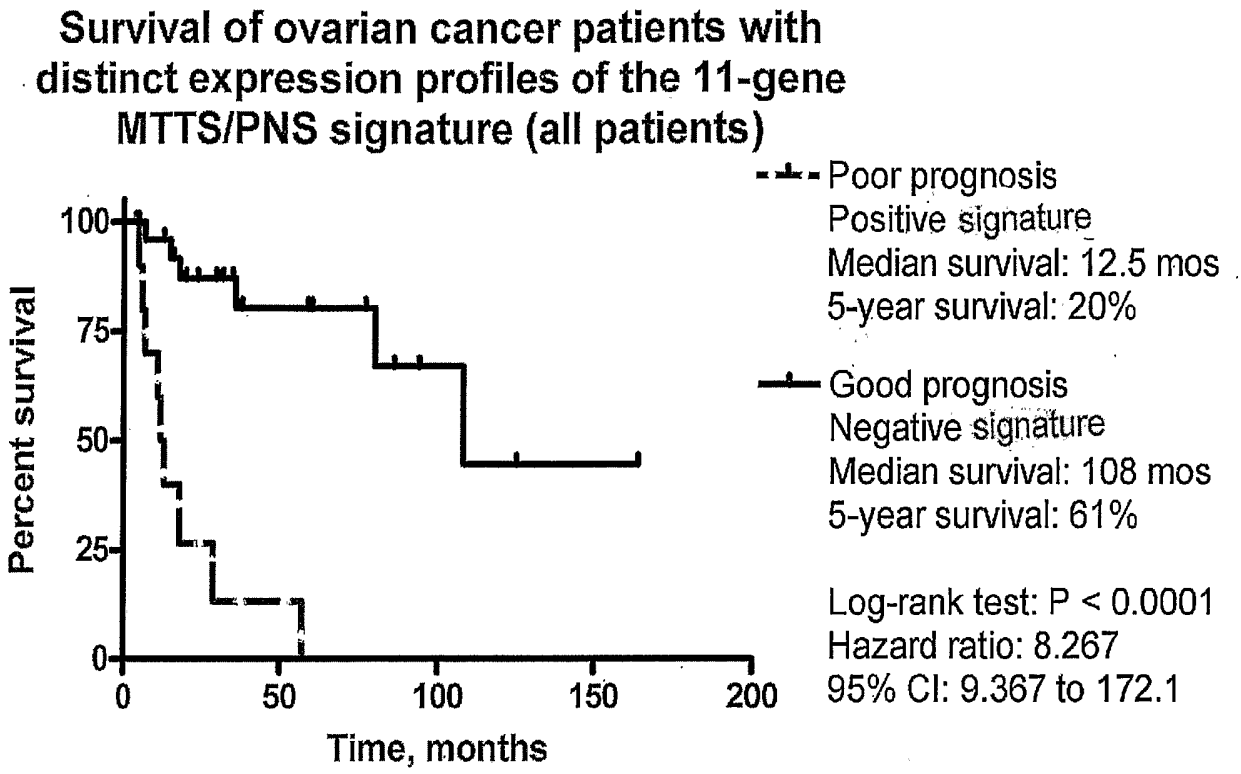


FIGURE 11

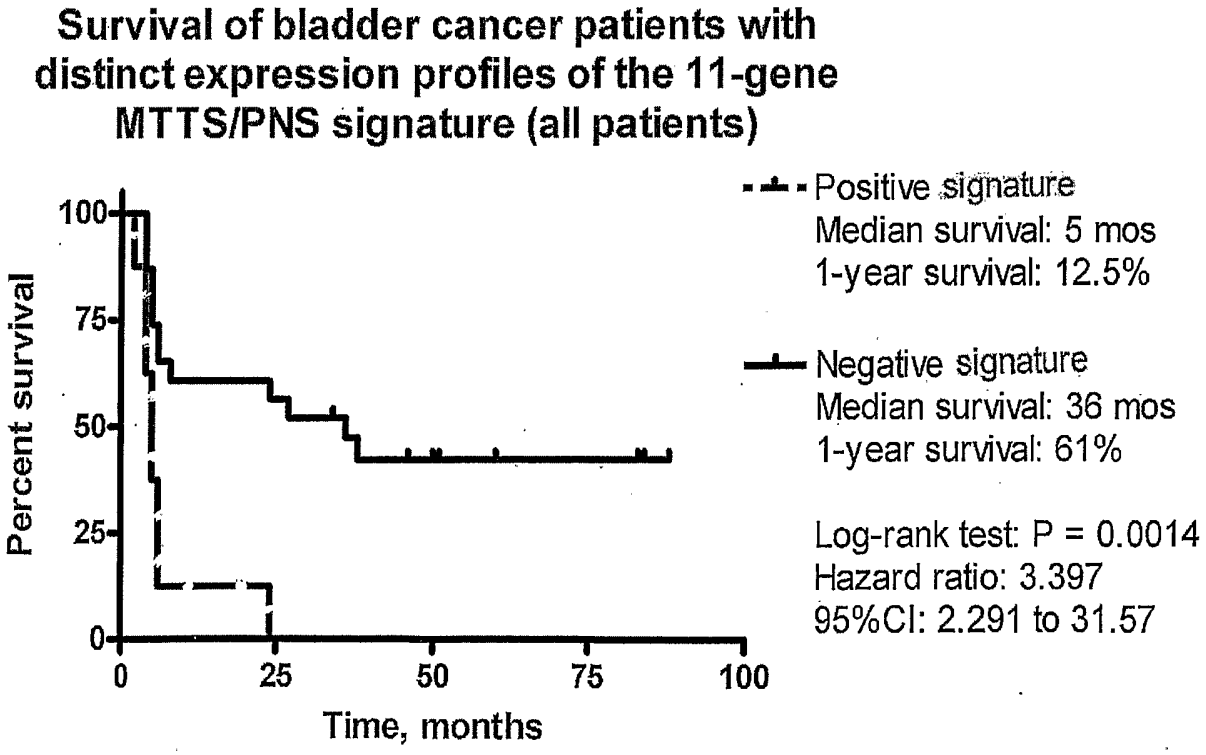


FIGURE 12

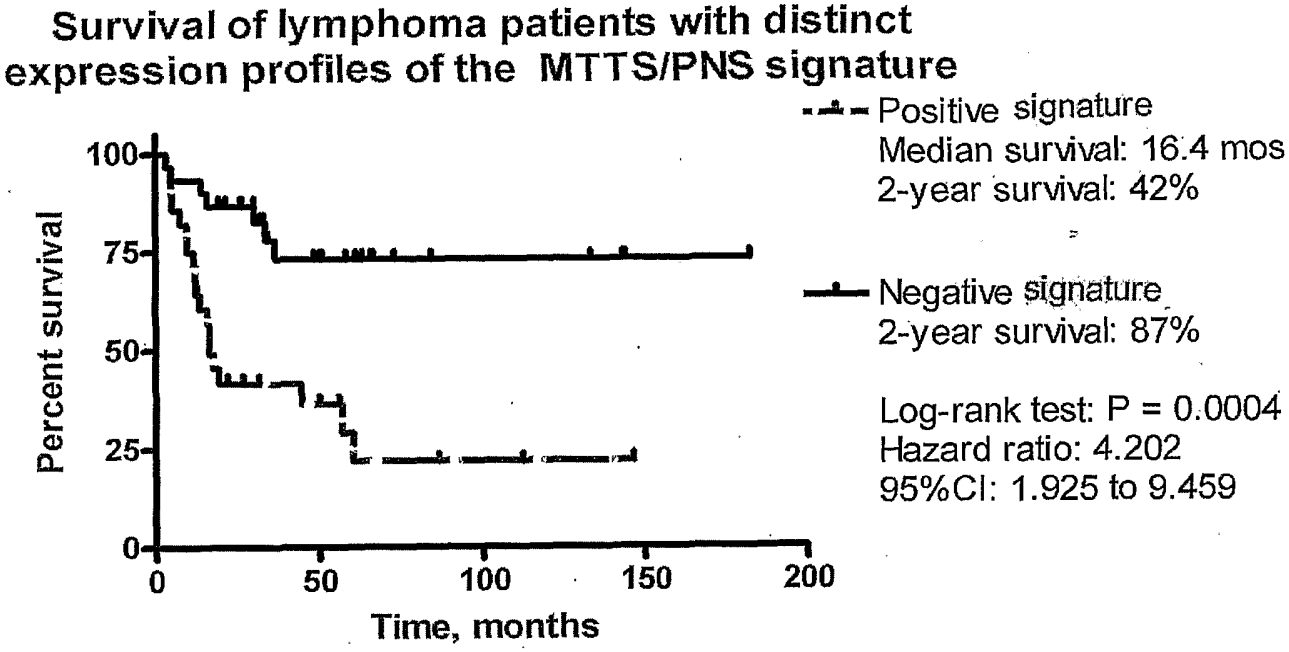


FIGURE 13

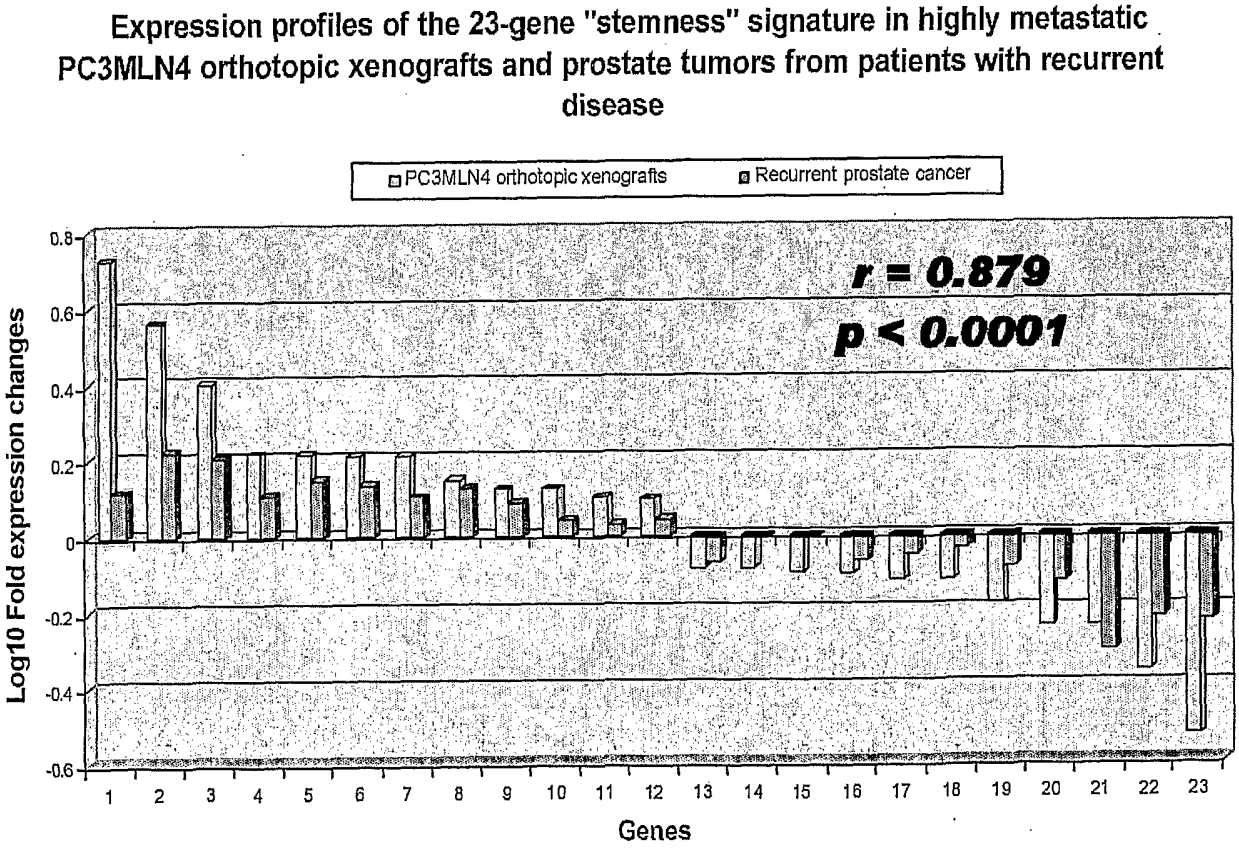


FIGURE 14

Expression profiles of the 16-gene "stemness" signature in distant prostate cancer metastases and primary prostate tumors from patients with recurrent disease

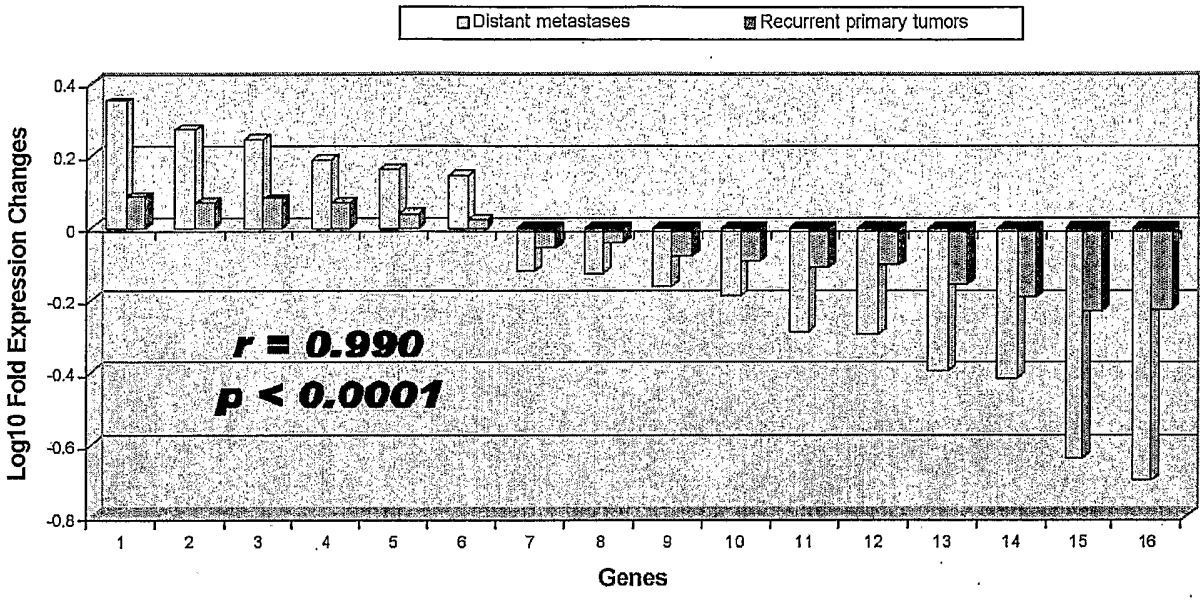


FIGURE 15

Expression profile of the 14-gene "stemness" signature in 8 recurrent versus 13 non-recurrent human prostate carcinomas

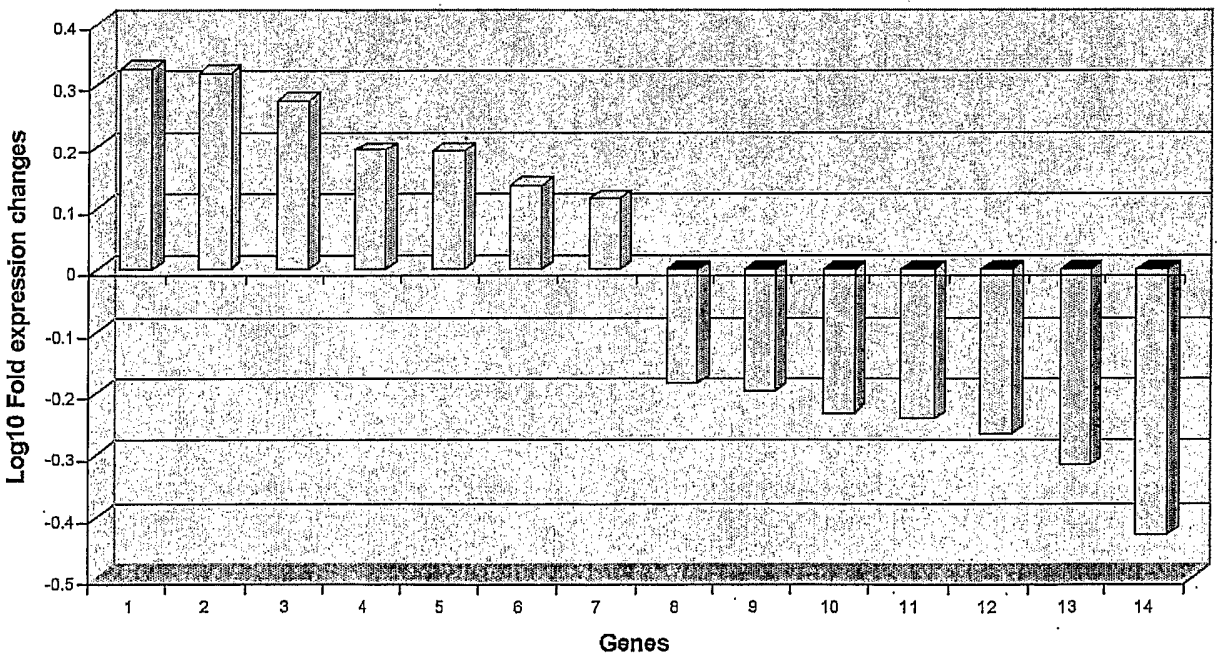


FIGURE 16

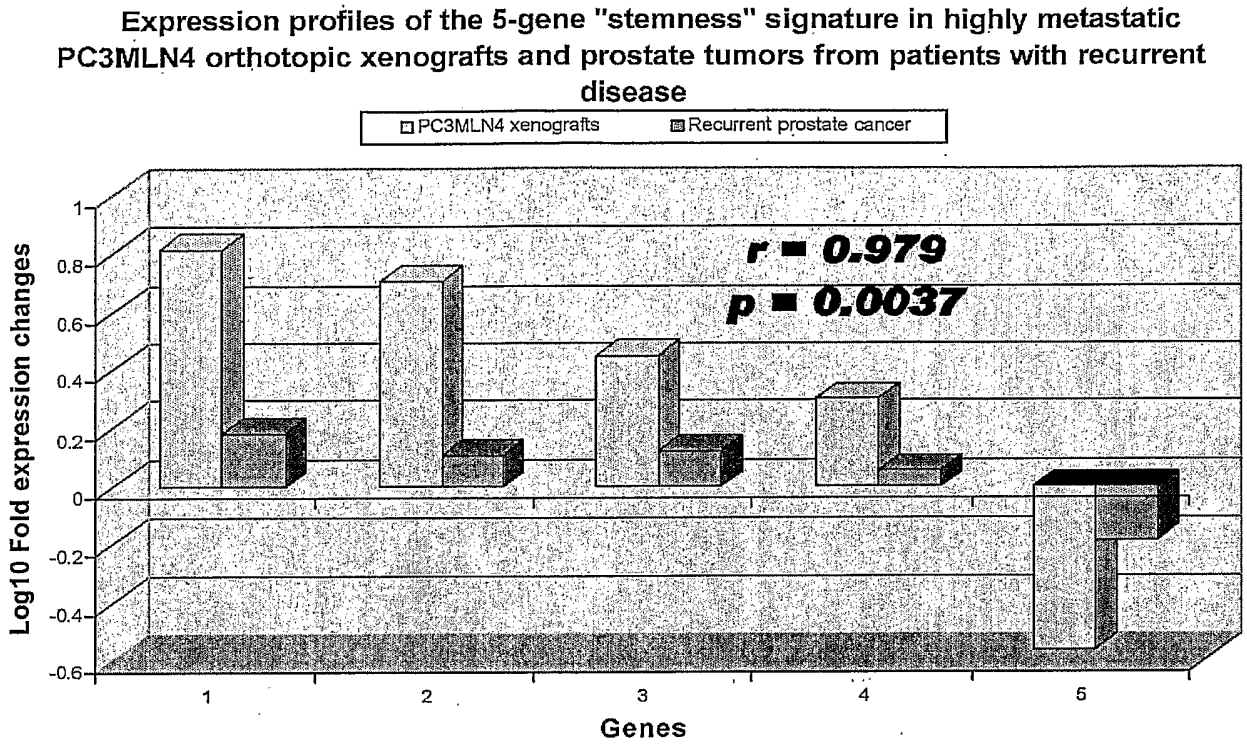


FIGURE 17

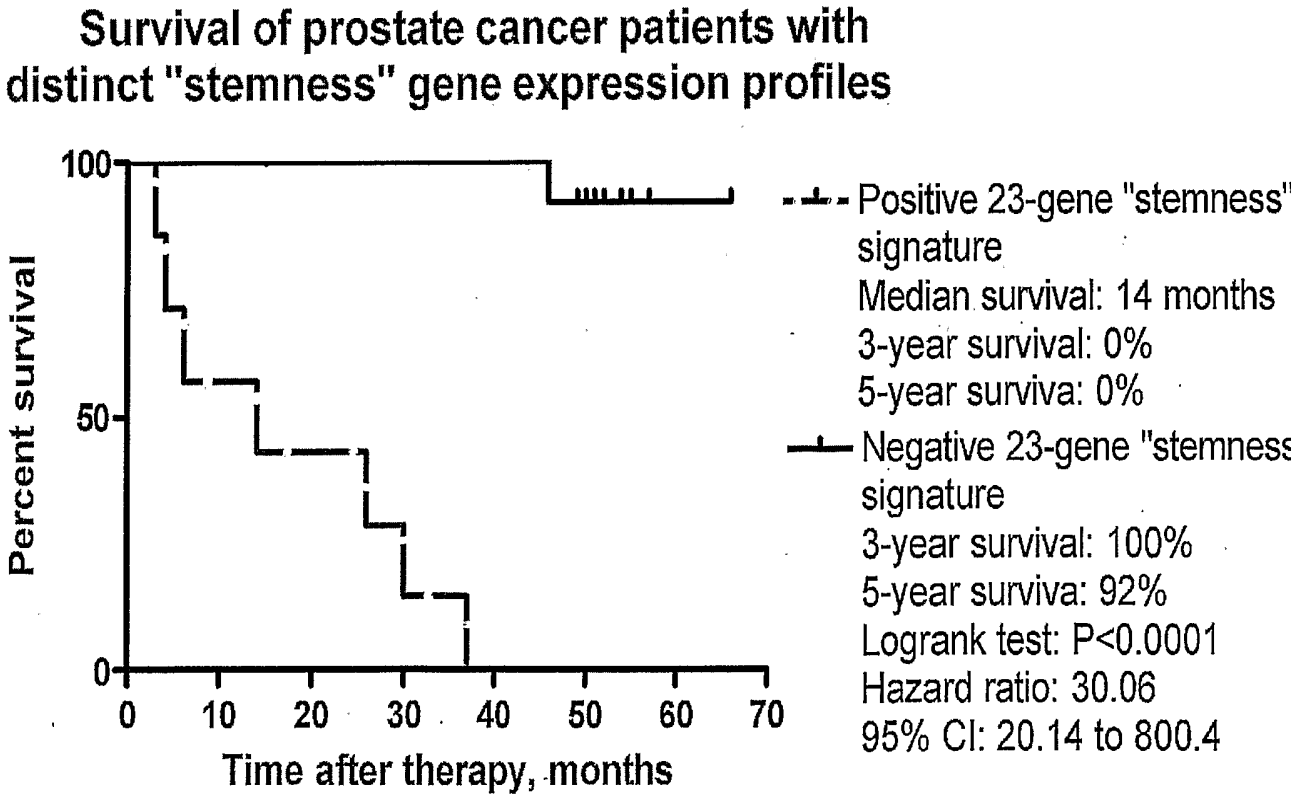


FIGURE 18

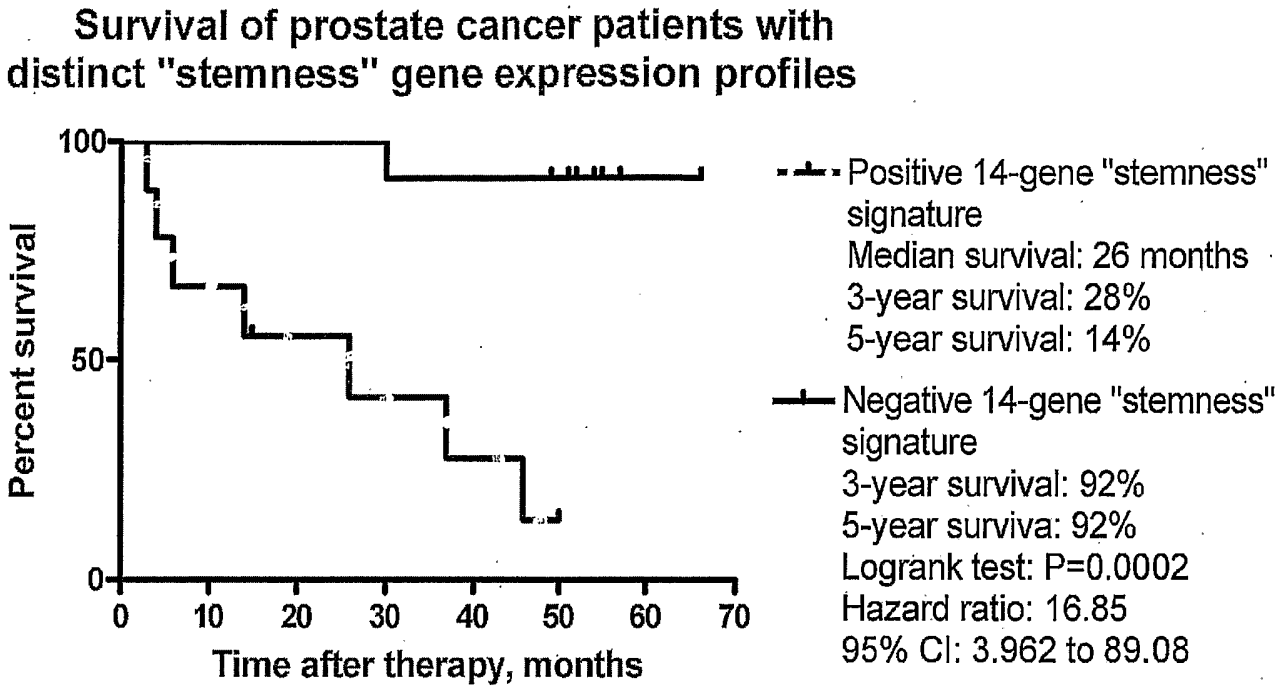


FIGURE 19

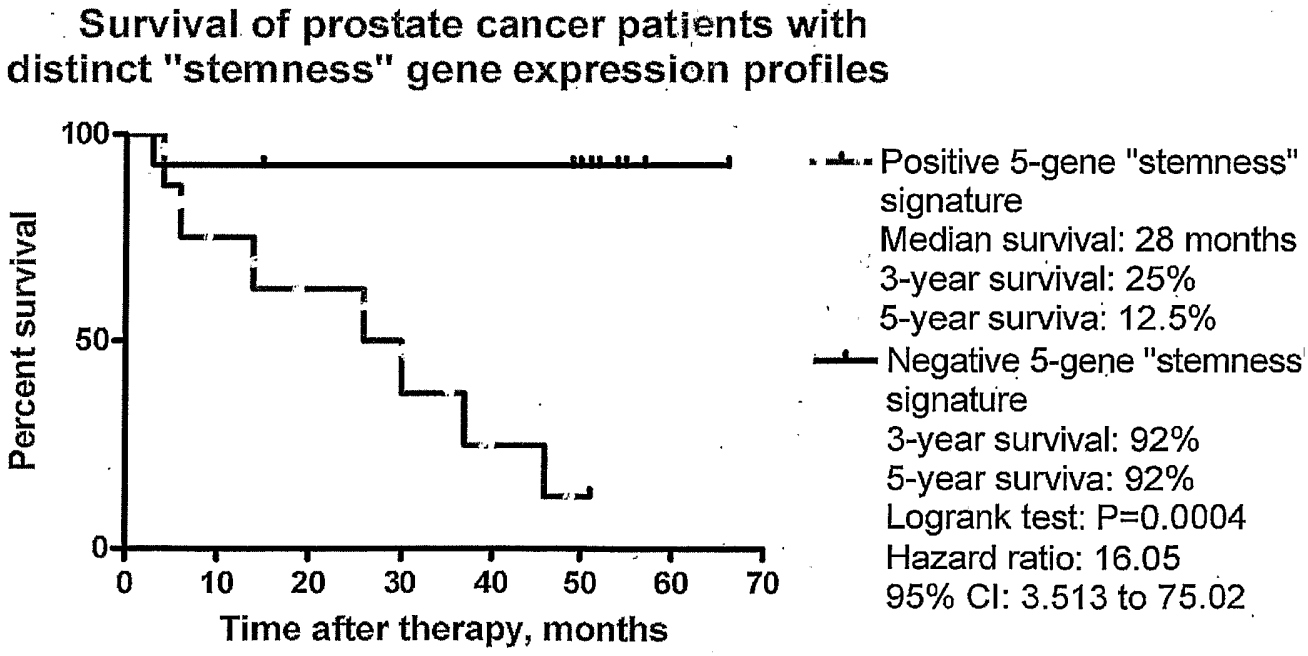


FIGURE 20

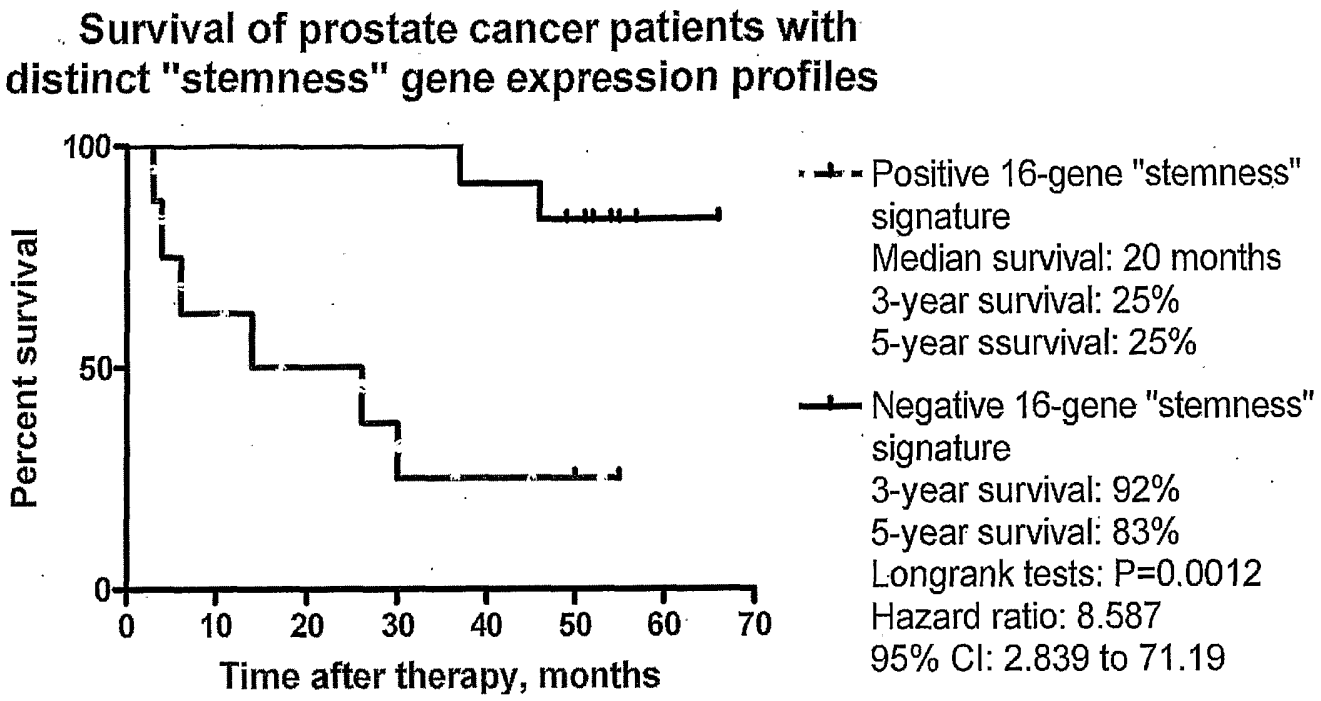


FIGURE 21

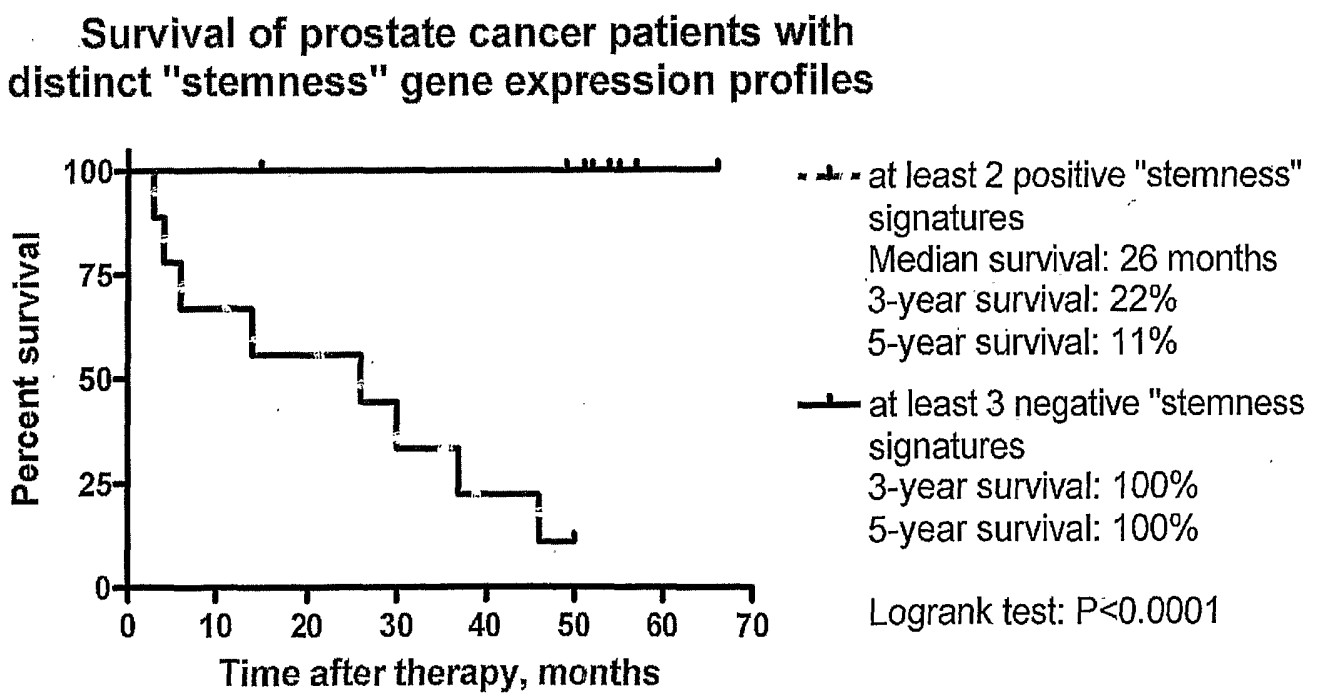


FIGURE 22

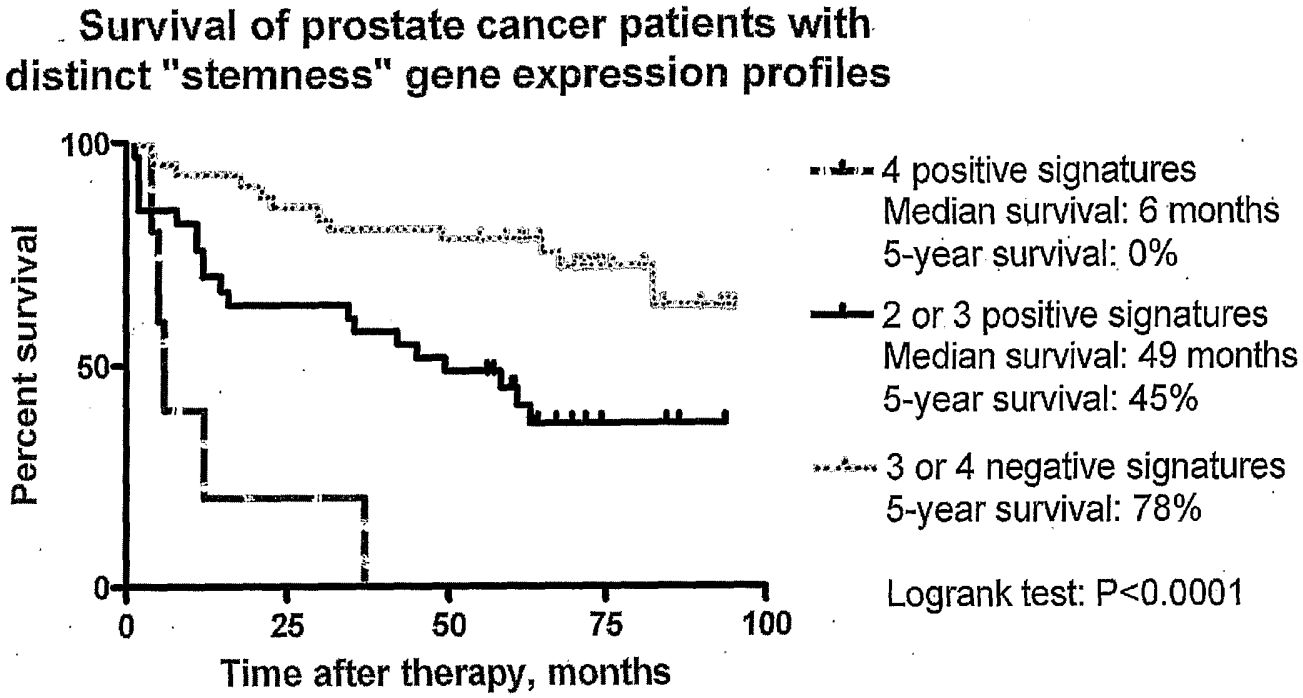


FIGURE 23

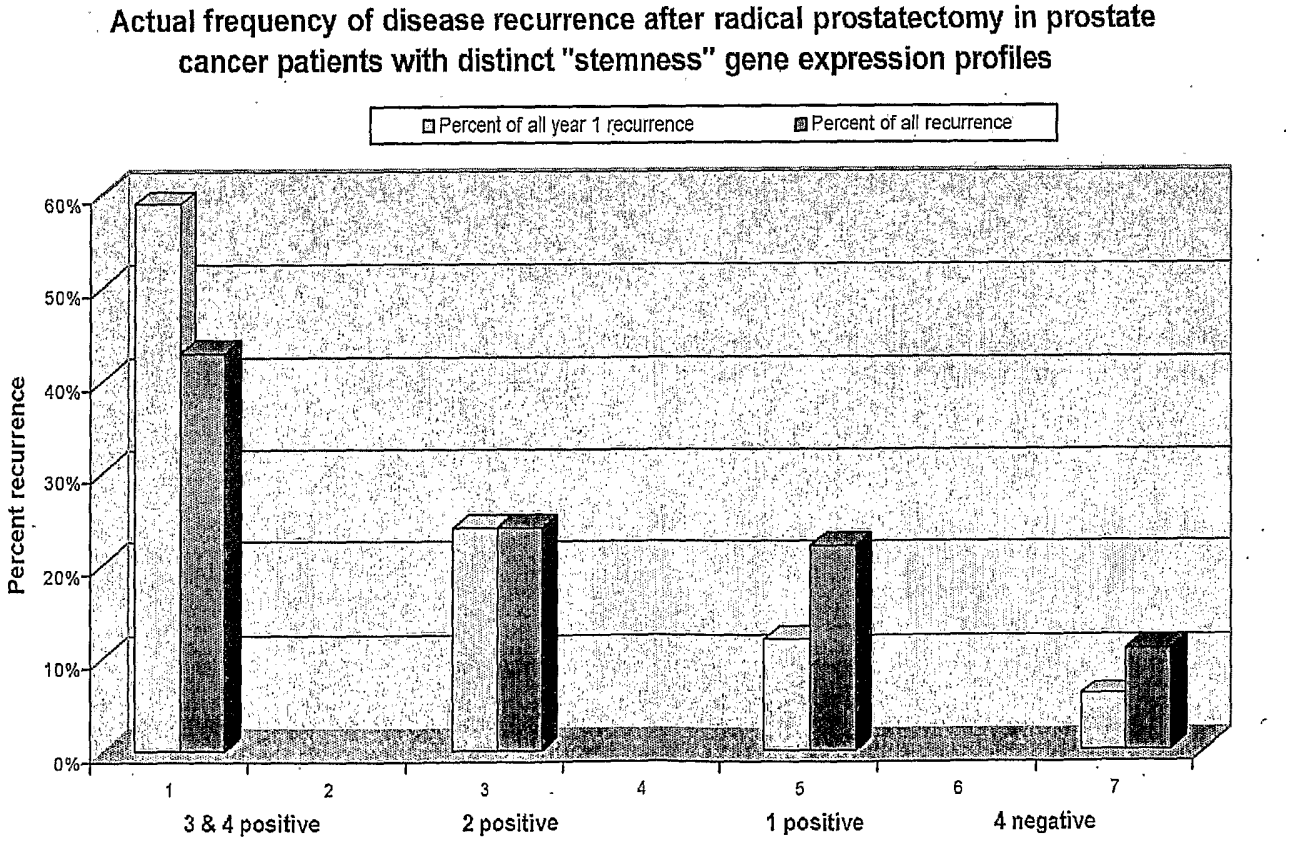


FIGURE 24

SEQUENCE LISTING

1. GBX2 (*Homo sapiens*) cDNA sequence

```
   1 atgagcgcag cgttcccgcc gtcgctgatg atgatgcagc gcccgctggg gagtagcacc
  61 gccttcagca tagactcgct gatcggcagc ccgccgcagc ccagccccgg ccatttcgtc
 121 tacaccggct accccatgtt catgccctac cggccggtag tgctgccgcc gccgccgccg
 181 ccgccgcccg cgctgcccca ggccgcgctg cagccagcgc tgccgccccgc acaccctcac
 241 caccagatcc ccagcctgcc cacaggcttc tgctccagcc tggcgcaggg catggcgctc
 301 acctctacgc tcatggccac gctccccggc ggcttctccg cgtcgcccca gcaccaggag
 361 gcggcagcgg cccgcaagtt cgcgccgcag ccgctgcccg gcggcggtaa cttcgacaag
 421 gcggaggcgc tgcaggctga cgcggaggac ggcaaaggct tcctggccaa agagggctcg
 481 ctgctcgcct tctccgcggc cgagacggtg caggcttcgc tcgtcggggc tgtccgaggg
 541 caagggaaag acgagtcaaa ggtggaagac gacccgaagg gcaaggagga gagcttctcg
 601 ctggagagcg atgtggacta cagctcggat gacaatctga ctggccaggc agctcacaag
 661 gaggaagacc cgggccacgc gctggaggag accccgccga gcagcggcgc cgcgggcagc
 721 accacgtcta cgggcaagaa ccggcggcgg cggactgcct tcaccagcga gcagctgctg
 781 gagctagaga aggagttcca ctgcaaaaag tacctctcct tgaccgagcg ctcgcagatc
 841 gcccacgccc tcaaactcag cgaggtgcag gtgaaaatct ggttccagaa ccgacgggcc
 901 aagtggaaac gggtgaaggc aggcaatgcc aattccaaga caggggagcc ctcccggaac
 961 cctaagatcg tcgtccccat ccctgtccac gtcagcaggt tcgctatcag aagtcagcat
1021 cagcagctag aacaggcccg gccctga
```

//

2. GBX2 (*Homo sapiens*) Protein sequence

```
   1 msaafppslm mmqrplgsst afsidsligs ppqpspghfv ytgypmfmpy rpvvlppppp
  61 pppalpqaal qpalppahph hqipslptgf csslaqgmal tstlmatlpg gfsaspqhqe
 121 aaaarkfapq plpgggnfdk aealqadaed gkgflakegs llafsaaetv qaslvgavrg
 181 qgkdeskved dpkgkeesfs lesdvdyssd dnltgqaahk eedpghalee tppssgaags
 241 ttstgknrrr rtaftseqll elekefhckk ylsltersqi ahalklsevq vkiwfqnrra
 301 kwkrvkagna nsktgepsrn pkivvpipvh vsrfairsqh qqleqarp
```

//

3. GBX2 (*Mus musculus*) cDNA sequence

```
   1 atgagcgcag cgttcccgcc gtcgctgatg atgatgcagc gcccgctggg gagtagtacc
  61 gccttcagca tagactcgct gatcggcagc ccgccgcagc ccagtcccgg ccatttcgtc
 121 tacaccggct accccatgtt catgccctac cggccggtgg tgctgccgcc accgccgcca
 181 ccgcctcccg cgctgcccca ggcagcgctg cagcccgctc tgccgcccgc gcaccctcac
 241 caccagatcc ccagcctgcc caccggcttc tgctccagcc tggcgcaggg catggcgctc
 301 acctccacgc tcatggccac tctgcccggc ggcttctctg cgtcgcccca gcaccaagag
 361 gcggcggctg cccgcaagtt cgctccacag ccactgcccg gaggcggcaa cttcgacaaa
 421 gccgaggcgc tccaagcgga tgcggaagac ggcaaagcct tcttggccaa ggagggctcg
 481 ctgctcgctt tctctgcggc cgaagcggtg caggcgtcgc tcgtcggggc tgtccgaggg
 541 caagggaaag acgagtcaaa ggtggaagat gacccgaagg gcaaggagga gagcttctct
 601 ctggagagcg atgtggatta cagctcagat gacaatttgc ctggtcagac tgctcataag
 661 gaagaagacc ccggccacgc actggaggag accccgcaga gcggcggtgc agcaggcagc
 721 accacgtcca caggcaagaa ccggcggcgg cggactgcct tcaccagcga acagctgctg
```

781 gagctggaga aagaattcca ctgcaaaaag tacctctccc tgaccgagcg ctcacagatc
841 gcccacgccc tcaaactcag cgaggtgcaa gtaaaaatct ggttccagaa ccgccgggcc
901 aagtggaaac gtgtcaaggc aggcaacgcc aattccaaga cgggggagcc ctctcggaac
961 cccaagattg tcgtccccat ccctgttcac gttagcaggt tcgctattcg aagtcaacac
1021 cagcagctgg agcaggcccg accctga

//

4. GBX2 (*Mus musculus*) Protein sequence

1 msaafppslm mmqrplgsst afsidsligs ppqpspghfv ytgypmfmpy rpvvlppppp
61 pppalpqaal qpalppahph hqipslptgf csslaqgmal tstlmatlpg gfsaspqhqe
121 aaaarkfapq plpgggnfdk aealqadaed gkaflakegs llafsaaeav qaslvgavrg
181 qgkdeskved dpkgkeesfs lesdvdyssd dnlpgqtahk eedpghalee tpqsggaags
241 ttstgknrrr rtaftseqll elekefhckk ylsltersqi ahalklsevq vkiwfqnrra
301 kwkrvkagna nsktgepsrn pkivvpipvh vsrfairsqh qqleqarp

//

5. MKI67 (*Homo sapiens*) cDNA sequence

1 atgtggccca cgagacgcct ggttactatc aaaaggagcg gggtcgacgg tccccacttt
61 cccctgagcc tcagcacctg cttgtttgga aggggtattg aatgtgacat ccgtatccag
121 cttcctgttg tgtcaaaaca acattgcaaa attgaaatcc atgagcagga ggcaatatta
181 cataatttca gttccacaaa tccaacacaa gtaaatgggt ctgttattga tgagcctgta
241 cggctaaaac atggagatgt aataactatt attgatcgtt ccttcaggta tgaaaatgaa
301 agtcttcaga atggaaggaa gtcaactgaa tttccaagaa aaatacgtga acaggagcca
361 gcacgtcgtg tctcaagatc tagcttctct tctgaccctg atgagaaagc tcaagattcc
421 aaggcctatt caaaaatcac tgaaggaaaa gtttcaggaa atcctcaggt acatatcaag
481 aatgtcaaag aagacagtac cgcagatgac tcaaaagaca gtgttgctca gggaacaact
541 aatgttcatt cctcagaaca tgctggacgt aatggcagaa atgcagctga tcccatttct
601 ggggatttta aagaaatttc cagcgttaaa ttagtgagcc gttatggaga attgaagtct
661 gttcccacta cacaatgtct tgacaatagc aaaaaaaatg aatctccctt ttggaagctt
721 tatgagtcag tgaagaaaga gttggatgta aaatcacaaa aagaaaatgt cctacagtat
781 tgtagaaaat ctggattaca aactgattac gcaacagaga aagaaagtgc tgatggttta
841 caggggggaga cccaactgtt ggtctcgcgt aagtcaagac caaaatctgg tgggagcggc
901 cacgctgtgg cagagcctgc ttcacctgaa caagagcttg accagaacaa ggggaaggga
961 agagacgtgg agtctgttca gactcccagc aaggctgtgg gcgccagctt tcctctctat
1021 gagccggcta aaatgaagac ccctgtacaa tattcacagc aacaaaattc tccacaaaaa
1081 cataagaaca aagacctgta tactactggt agaagagaat ctgtgaatct gggtaaaagt
1141 gaaggcttca aggctggtga taaaactctt actcccagga agctttcaac tagaaatcga
1201 acaccagcta aagttgaaga tgcagctgac tctgccacta agccagaaaa tctctcttcc
1261 aaaaccagag gaagtattcc tacagatgtg gaagttctgc ctacggaaac tgaaattcac
1321 aatgagccat ttttaactct gtggctcact caagttgaga ggaagatcca aaaggattcc
1381 ctcagcaagc ctgagaaatt gggcactaca gctggacaga tgtgctctgg gttacctggt
1441 cttagttcag ttgatatcaa caactttggt gattccatta atgagagtga gggaatacct
1501 ttgaaaagaa ggcgtgtgtc ctttggtggg cacctaagac ctgaactatt tgatgaaaac
1561 ttgcctccta atacgcctct caaaagggga gaagccccaa ccaaaagaaa gtctctggta
1621 atgcacactc cacctgtcct gaagaaaatc atcaaggaac agcctcaacc atcaggaaaa
1681 caagagtcag gttcagaaat ccatgtggaa gtgaaggcac aaagcttggt tataagccct

1741 ccagctccta gtcctaggaa aactccagtt gccagtgatc aacgccgtag gtcctgcaaa
1801 acagcccctg cttccagcag caaatctcag acagaggttc ctaagagagg aggagaaaga
1861 gtggcaacct gccttcaaaa gagagtgtct atcagccgaa gtcaacatga tattttacag
1921 atgatatgtt ccaaaagaag aagtggtgct tcggaagcaa atctgattgt tgcaaaatca
1981 tgggcagatg tagtaaaact tggtgcaaaa caaacacaaa ctaaagtcat aaaacatggt
2041 cctcaaaggt caatgaacaa aaggcaaaga agacctgcta ctccaaagaa gcctgtgggc
2101 gaagttcaca gtcaatttag tacaggccac gcaaactctc cttgtaccat aataataggg
2161 aaagctcata ctgaaaaagt acatgtgcct gctcgaccct acagagtgct caacaacttc
2221 atttccaacc aaaaaatgga ctttaaggaa gatctttcag gaatagctga aatgttcaag
2281 accccagtga aggagcaacc gcagttgaca agcacatgtc acatcgctat ttcaaattca
2341 gagaatttgc ttggaaaaca gtttcaagga actgattcag gagaagaacc tctgctcccc
2401 acctcagaga gttttggagg aaatgtgttc ttcagtgcac agaatgcagc aaaacagcca
2461 tctgataaat gctctgcaag ccctccctta agacggcagt gtattagaga aaatggaaac
2521 gtagcaaaaa cgcccaggaa cacctacaaa atgacttctc tggagacaaa aacttcagat
2581 actgagacag agccttcaaa aacagtatcc actgtaaaca ggtcaggaag gtctacagag
2641 ttcaggaata tacagaagct acctgtggaa agtaagagtg aagaaacaaa tacagaaatt
2701 gttgagtgca tcctaaaaag aggtcagaag gcaacactac tacaacaaag gagagaagga
2761 gagatgaagg aaatagaaag acctttgag acatataagg aaaatattga attaaaagaa
2821 aacgatgaaa agatgaaagc aatgaagaga tcaagaactt gggggcagaa atgtgcacca
2881 atgtctgacc tgacagacct caagagcttg cctgatacag aactcatgaa agacacggca
2941 cgtggccaga atctcctcca aacccaagat catgccaagg caccaaagag tgagaaaggc
3001 aaaatcacta aaatgccctg ccagtcatta caaccagaac caataaacac cccaacacac
3061 acaaaacaac agttgaaggc atccctgggg aaagtaggtg tgaaagaaga gctcctagca
3121 gtcggcaagt tcacacggac gtcaggggag accacgcaca cgcacagaga gccagcagga
3181 gatggcaaga gcatcagaac gtttaaggag tctccaaagc agatcctgga cccagcagcc
3241 cgtgtaactg gaatgaagaa gtggccaaga acgcctaagg aagaggccca gtcactagaa
3301 gacctggctg gcttcaaaga gctcttccag acaccaggtc cctctgagga atcaatgact
3361 gatgagaaaa ctaccaaaat agcctgcaaa tctccaccac cagaatcagt ggacactcca
3421 acaagcacaa agcaatggcc taagagaagt ctcaggaaag cagatgtaga ggaagaattc
3481 ttagcactca ggaaactaac accatcagca gggaaagcca tgcttacgcc caaaccagca
3541 ggaggtgatg agaaagacat taaagcattt atgggaactc cagtgcagaa actggacctg
3601 gcaggaactt tacctggcag caaaagacag ctacagactc ctaaggaaaa ggcccaggct
3661 ctagaagacc tggctggctt taaagagctc ttccagactc ctggtcacac cgaggaatta
3721 gtggctgctg gtaaaaccac taaaataccc tgcgactctc cacagtcaga cccagtggac
3781 accccaacaa gcacaaagca acgacccaag agaagtatca ggaaagcaga tgtagaggga
3841 gaactcttag cgtgcaggaa tctaatgcca tcagcaggca aagccatgca cacgcctaaa
3901 ccatcagtag gtgaagagaa agacatcatc atatttgtgg gaactccagt gcagaaactg
3961 gacctgacag agaacttaac cggcagcaag agacggccac aaactcctaa ggaagaggcc
4021 caggctctgg aagacctgac tggctttaaa gagctcttcc agacccctgg tcatactgaa
4081 gaagcagtgg ctgctggcaa aactactaaa atgccctgcg aatcttctcc accagaatca
4141 gcagacaccc caacaagcac aagaaggcag cccaagacac ctttggagaa aagggacgta
4201 cagaaggagc tctcagccct gaagaagctc acacagacat caggggaaac cacacacaca
4261 gataaagtac caggaggtga ggataaaagc atcaacgcgt ttagggaaac tgcaaaacag
4321 aaactggacc cagcagcaag tgtaactggt agcaagaggc acccaaaaac taaggaaaag

4381 gcccaacccc tagaagacct ggctggctgg aaagagctct tccagacacc agtatgcact
4441 gacaagccca cgactcacga gaaaactacc aaaatagcct gcagatcaca accagaccca
4501 gtggacacac caacaagctc caagccacag tccaagagaa gtctcaggaa agtggacgta
4561 gaagaagaat tcttcgcact caggaaacga acaccatcag caggcaaagc catgcacaca
4621 cccaaaccag cagtaagtgg tgagaaaaac atctacgcat ttatgggaac tccagtgcag
4681 aaactggacc tgacagagaa cttaactggc agcaagagac ggctacaaac tcctaaggaa
4741 aaggcccagg ctctagaaga cctggctggc tttaaagagc tcttccagac acgaggtcac
4801 actgaggaat caatgactaa cgataaaact gccaaagtag cctgcaaatc ttcacaacca
4861 gacctagaca aaaacccagc aagctccaag cgacggctca agacatccct ggggaaagtg
4921 ggcgtgaaag aagagctcct agcagttggc aagctcacac agacatcagg agagactaca
4981 cacacacaca cagagccaac aggagatggt aagagcatga aagcatttat ggagtctcca
5041 aagcagatct tagactcagc agcaagtcta actggcagca agaggcagct gagaactcct
5101 aagggaaagt ctgaagtccc tgaagacctg gccggcttca tcgagctctt ccagacacca
5161 agtcacacta aggaatcaat gactaatgaa aaaactacca aagtatccta cagagcttca
5221 cagccagacc tagtggacac cccaacaagc tccaagccac agcccaagag aagtctcagg
5281 aaagcagaca ctgaagaaga atttttagca tttaggaaac aaacgccatc agcaggcaaa
5341 gccatgcaca cacccaaacc agcagtaggt gaagagaaag acatcaacac gtttttggga
5401 actccagtgc agaaactgga ccagccagga aatttacctg gcagcaatag acggctacaa
5461 actcgtaagg aaaaggccca ggctctagaa gaactgactg gcttcagaga gcttttccag
5521 acaccatgca ctgataaccc cacagctgat gagaaaacta ccaaaaaaat actctgcaaa
5581 tctccgcaat cagacccagc ggacacccca acaaacacaa agcaacggcc caagagaagc
5641 ctcaagaaag cagacgtaga ggaagaattt ttagcattca ggaaactaac accatcagca
5701 ggcaaagcca tgcacacgcc taaagcagca gtaggtgaag agaaagacat caacacattt
5761 gtggggactc cagtggagaa actggacctg ctaggaaatt tacctggcag caagagacgg
5821 ccacaaactc ctaaagaaaa ggccaaggct ctagaagatc tggctggctt caaagagctc
5881 ttccagacac caggtcacac tgaggaatca atgaccgatg acaaaatcac agaagtatcc
5941 tgcaaatctc cacaaccaga cccagtcaaa accccaacaa gctccaagca acgactcaag
6001 atatccttgg ggaaagtagg tgtgaaagaa gaggtcctac cagtcggcaa gctcacacag
6061 acgtcaggga agaccacaca gacacacaga gagacagcag gagatggaaa gagcatcaaa
6121 gcgtttaagg aatctgcaaa gcagatgctg gacccagcaa actatggaac tgggatggag
6181 aggtggccaa gaacacctaa ggaagaggcc caatcactag aagacctggc cggcttcaaa
6241 gagctcttcc agacaccaga ccacactgag gaatcaacaa ctgatgacaa aactaccaaa
6301 atagcctgca aatctccacc accagaatca atggacactc caacaagcac aaggaggcgg
6361 cccaaaacac ctttggggaa aagggatata gtggaagagc tctcagccct gaagcagctc
6421 acacagacca cacacacaga caaagtacca ggagatgagg ataaaggcat caacgtgttc
6481 agggaaactg caaaacagaa actggaccca gcagcaagtg taactggtag caagaggcag
6541 ccaagaactc ctaagggaaa agcccaaccc ctagaagact tggctggctt gaaagagctc
6601 ttccagacac cagtatgcac tgacaagccc acgactcacg agaaaactac caaaatagcc
6661 tgcagatctc cacaaccaga cccagtgggt accccaacaa tcttcaagcc acagtccaag
6721 agaagtctca ggaaagcaga cgtagaggaa gaatccttag cactcaggaa acgaacacca
6781 tcagtaggga aagctatgga cacacccaaa ccagcaggag gtgatgagaa agacatgaaa
6841 gcatttatgg gaactccagt gcagaaattg gacctgccag gaaatttacc tggcagcaaa
6901 agatggccac aaactcctaa ggaaaaggcc caggctctag aagacctggc tggcttcaaa
6961 gagctcttcc agacaccagg cactgacaag cccacgactg atgagaaaac taccaaaata

7021 gcctgcaaat ctccacaacc agacccagtg gacaccccag caagcacaaa gcaacggccc
7081 aagagaaacc tcaggaaagc agacgtagag gaagaatttt tagcactcag gaaacgaaca
7141 ccatcagcag gcaaagccat ggacacccca aaaccagcag taagtgatga gaaaaatatc
7201 aacacatttg tggaaactcc agtgcagaaa ctggacctgc taggaaattt acctggcagc
7261 aagagacagc cacagactcc taaggaaaag gctgaggctc tagaggacct ggttggcttc
7321 aaagaactct tccagacacc aggtcacact gaggaatcaa tgactgatga caaaatcaca
7381 gaagtatcct gtaaatctcc acagccagag tcattcaaaa cctcaagaag ctccaagcaa
7441 aggctcaaga tacccctggt gaaagtggac atgaaagaag agcccctagc agtcagcaag
7501 ctcacacgga catcagggga gactacgcaa acacacacag agccaacagg agatagtaag
7561 agcatcaaag cgtttaagga gtctccaaag cagatcctgg acccagcagc aagtgtaact
7621 ggtagcagga ggcagctgag aactcgtaag gaaaaggccc gtgctctaga agacctggtt
7681 gacttcaaag agctcttctc agcaccaggt cacactgaag agtcaatgac tattgacaaa
7741 aacacaaaaa ttccctgcaa atctcccccа ccagaactaa cagacactgc cacgagcaca
7801 aagagatgcc ccaagacacg tcccaggaaa gaagtaaaag aggagctctc agcagttgag
7861 aggctcacgc aaacatcagg gcaaagcaca cacacacaca aagaaccagc aagcggtgat
7921 gagggcatca aagtattgaa gcaacgtgca aagaagaaac caaacccagt agaagaggaa
7981 cccagcagga gaaggccaag agcacctaag gaaaaggccc aacccctgga agacctggcc
8041 ggcttcacag agctctctga aacatcaggt cacactcagg aatcactgac tgctggcaaa
8101 gccactaaaa taccctgcga atctcccccа ctagaagtgg tagacaccac agcaagcaca
8161 aagaggcatc tcaggacacg tgtgcagaag gtacaagtaa aagaagagcc ttcagcagtc
8221 aagttcacac aaacatcagg ggaaaccacg gatgcagaca aagaaccagc aggtgaagat
8281 aaaggcatca aagcattgaa ggaatctgca aaacagacac cggctccagc agcaagtgta
8341 actggcagca ggagacggcc aagagcaccc agggaaagtg cccaagccat agaagaccta
8401 gctggcttca aagacccagc agcaggtcac actgaagaat caatgactga tgacaaaacc
8461 actaaaatac cctgcaaatc atcaccagaa ctagaagaca ccgcaacaag ctcaaagaga
8521 cggcccagga cacgtgccca gaaagtagaa gtgaaggagg agctgttagc agttggcaag
8581 ctcacacaaa cctcagggga gaccacgcac accgacaaag agccggtagg tgagggcaaa
8641 ggcacgaaag catttaagca acctgcaaag cggaacgtgg acgcagaaga tgtaattggc
8701 agcaggagac agccaagagc acctaaggaa aaggcccaac ccctggaaga cctggccagc
8761 ttccaagagc tctctcaaac accaggccac actgaggaac tggcaaatgg tgctgctgat
8821 agctttacaa gcgctccaaa gcaaacacct gacagtggaa aacctctaaa aatatccaga
8881 agagttcttc gggcccctaa agtagaaccc gtgggagacg tggtaagcac cagagaccct
8941 gtaaaatcac aaagcaaaag caacacttcc ctgcccccac tgcccttcaa gaggggaggt
9001 ggcaaagatg gaagcgtcac gggaaccaag aggctgcgct gcatgccagc accagaggaa
9061 attgtggagg agctgccagc cagcaagaag cagagggttg ctcccagggc aagaggcaaa
9121 tcatccgaac ccgtggtcat catgaagaga agtttgagga cttctgcaaa aagaattgaa
9181 cctgcggaag agctgaacag caacgacatg aaaaccaaca aagaggaaca caaattacaa
9241 gactcggtcc ctgaaaataa gggaatatcc ctgcgctcca gacgccaaga taagactgag
9301 gcagaacagc aaataactga ggtctttgta ttagcagaaa gaatagaaat aaacagaaat
9361 gaaaagaagc ccatgaagac ctccccagag atggacattc agaatccaga tgatggagcc
9421 cggaaaccca tacctagaga caaagtcact gagaacaaaa ggtgcttgag gtctgctaga
9481 cagaatgaga gctcccagcc taaggtggca gaggagagcg gagggcagaa gagtgcgaag
9541 gttctcatgc agaatcagaa agggaaagga gaagcaggaa attcagactc catgtgcctg
9601 agatcaagaa agacaaaaag ccagcctgca gcaagcactt tggagagcaa atctgtgcag

```
 9661 agagtaacgc ggagtgtcaa gaggtgtgca gaaaatccaa agaaggctga ggacaatgtg
 9721 tgtgtcaaga aaataacaac cagaagtcat agggacagtg aagatatttg a
//
```

6. MKI67 (*Homo sapiens*) Protein sequence

```
    1 mwptrrlvti krsgvdgphf plslstclfg rgiecdiriq lpvvskqhck ieiheqeail
   61 hnfsstnptq vngsvidepv rlkhgdviti idrsfryene slqngrkste fprkireqep
  121 arrvsrssfs sdpdekaqds kayskitegk vsgnpqvhik nvkedstadd skdsvaqgtt
  181 nvhssehagr ngrnaadpis gdfkeissvk lvsrygelks vpttqcldns kknespfwkl
  241 yesvkkeldv ksqkenvlqy crksglqtdy atekesadgl qgetqllvsr ksrpksggsg
  301 havaepaspe qeldqnkgkg rdvesvqtps kavgasfply epakmktpvq ysqqqnspqk
  361 hknkdlyttg rresvnlgks egfkagdktl tprklstrnr tpakvedaad satkpenlss
  421 ktrgsiptdv evlpteteih nepfltlwlt qverkiqkds lskpeklgtt agqmcsglpg
  481 lssvdinnfg dsinesegip lkrrrvsfgg hlrpelfden lppntplkrg eaptkrkslv
  541 mhtppvlkki ikeqpqpsgk qesgseihve vkaqslvisp papsprktpv asdqrrrsck
  601 tapasssksq tevpkrgger vatclqkrvs isrsqhdilq micskrrsga seanlivaks
  661 wadvvklgak qtqtkvikhg pqrsmnkrqr rpatpkkpvg evhsqfstgh anspctiiig
  721 kahtekvhvp arpyrvlnnf isnqkmdfke dlsgiaemfk tpvkeqpqlt stchiaisns
  781 enllgkqfqg tdsgeepllp tsesfggnvf fsaqnaakqp sdkcsasppl rrqcirengn
  841 vaktprntyk mtsletktsd tetepsktvs tvnrsgrste frniqklpve skseetntei
  901 vecilkrgqk atllqqrreg emkeierpfe tykenielke ndekmkamkr srtwgqkcap
  961 msdltdlks1 pdtelmkdta rgqnllqtqd hakapksekg kitkmpcqsl qpepintpth
 1021 tkqqlkaslg kvgvkeella vgkftrtsge tththrepag dgksirtfke spkqildpaa
 1081 rvtgmkkwpr tpkeeaqsle dlagfkelfq tpgpseesmt dekttkiack spppesvdtp
 1141 tstkqwpkrs lrkadveeef lalrkltpsa gkamltpkpa ggdekdikaf mgtpvqkldl
 1201 agtlpgskrq lqtpkekaqa ledlagfke1 fqtpghteel vaagkttkip cdspqsdpvd
 1261 tptstkqrpk rsirkadveg ellacrnlmp sagkamhtpk psvgeekdii ifvgtpvqkl
 1321 dltenltgsk rrpqtpkeea qaledltgfk elfqtpghte eavaagkttk mpcessppes
 1381 adtptstrrq pktplekrdv qkelsalkkl tqtsgettht dkvpggedks inafretakq
 1441 kldpaasvtg skrhpktkek aqpledlagw kelfqtpvct dkptthektt kiacrsqpdp
 1501 vdtptsskpq skrslrkvdv eeeffalrkr tpsagkamht pkpavsgekn iyafmgtpvq
 1561 kldltenltg skrrlqtpke kaqaledlag fkelfqtrgh teesmtndkt akvackssqp
 1621 dldknpassk rrlktslgkv gvkeellavg kltqtsgett hthteptgdg ksmkafmesp
 1681 kqildsaasl tgskrqlrtp kgksevpedl agfielfqtp shtkesmtne kttkvsyras
 1741 qpdlvdtpts skpqpkrslr kadteeefla frkqtpsagk amhtpkpavg eekdintflg
 1801 tpvqkldqpg nlpgsnrrlq trkekaqale eltgfrelfq tpctdnptad ekttkkilck
 1861 spqsdpadtp tntkqrpkrs lkkadveeef lafrkltpsa gkamhtpkaa vgeekdintf
 1921 vgtpvekldl lgnlpgskrr pqtpkekaka ledlagfkel fqtpghtees mtddkitevs
 1981 ckspqpdpvk tptsskqrlk islgkvgvke evlpvgkltq tsgkttqthr etagdgksik
 2041 afkesakqml dpanygtgme rwprtpkeea qsledlagfk elfqtpdhte esttddkttk
 2101 iackspppes mdtptstrrr pktplgkrdi veelsalkql tqtthtdkvp gdedkginvf
 2161 retakqkldp aasvtgskrq prtpkgkaqp ledlaglkel fqtpvctdkp tthekttkia
 2221 crspqpdpvg tptifkpqsk rslrkadvee eslalrkrtp svgkamdtpk paggdekdmk
 2281 afmgtpvqkl dlpgnlpgsk rwpqtpkeka qaledlagfk elfqtpgtdk pttdekttki
 2341 ackspqpdpv dtpastkqrp krnlrkadve eeflalrkrt psagkamdtp kpavsdekni
```

2401 ntfvetpvqk ldllgnlpgs krqpqtpkek aealedlvgf kelfqtpght eesmtddkit
2461 evsckspqpe sfktsrsskq rlkiplvkvd mkeeplavsk ltrtsgettq thteptgdsk
2521 sikafkespk qildpaasvt gsrrqlrtrk ekaraledlv dfkelfsapg hteesmtidk
2581 ntkipckspp peltdtatst krcpktrprk evkeelsave rltqtsgqst hthkepasgd
2641 egikvlkqra kkkpnpveee psrrrprapk ekaqpledla gftelsetsg htqesltagk
2701 atkipcespp levvdttast krhlrtrvqk vqvkeepsav kftqtsgett dadkepaged
2761 kgikalkesa kqtpapaasv tgsrrrprap resaqaiedl agfkdpaagh teesmtddkt
2821 tkipcksspe ledtatsskr rprtraqkve vkeellavgk ltqtsgetth tdkepvgegk
2881 gtkafkqpak rnvdaedvig srrqprapke kaqpledlas fqelsqtpgh teelangaad
2941 sftsapkqtp dsgkplkisr rvlrapkvep vgdvvstrdp vksqsksnts lpplpfkrgg
3001 gkdgsvtgtk rlrcmpapee iveelpaskk qrvaprargk ssepvvimkr slrtsakrie
3061 paeelnsndm ktnkeehklq dsvpenkgis lrsrrqdkte aeqqitevfv laerieinrn
3121 ekkpmktspe mdiqnpddga rkpiprdkvt enkrclrsar qnessqpkva eesgggksak
3181 vlmqnqkgkg eagnsdsmcl rsrktksqpa astlesksvq rvtrsvkrca enpkkaednv
3241 cvkkittrsh rdsedi

//

7. MKI67 (*Mus musculus*) cDNA sequence

1 atggcgtcct cggctcacct ggtcaccatc aagcggagcg gcgatgacgg cgcacacttc
61 ccgctgagcc tcagctcctg cctgtttgga aggagtattg aatgtgacat tcgtatccag
121 ctgcctgtag tgtctcaaag acattgccca attgtagtcc aagagcaaga ggcgatatta
181 tataatttca gttctaccaa tccaactcaa gtaaacgggg ttactataga tgagcctgtg
241 aggctgagac atggagacat aataaccatc attgaccgct cctttaggta tgaagatgga
301 aatcatgagg atggaagcaa accaacagaa tttccaggaa agtcccttgg aaaggaacca
361 tcaaggcgag cctcaagaga tagcttctgt gctgaccctg atggggaagg tcaagatacc
421 aaagcttcaa aaatgactgc ttcaagaaga tcttttgtgt atgccaaggg cctttctgca
481 gatagccctg cctcagatgg ctcaaagaac agtgttagcc aagactcatc agggcatgta
541 gaacagcaca ctggcagaaa catagtagag cccacttctg ggggatctct tttaagaagt
601 ccaggtctac agggagcagt tacagggaac cgaagtcttc ttcctacaca gagccttagc
661 aatagcaacg aaaaggaatc tccctttgag aaactttatc aatcaatgaa ggaagagttg
721 gatgtaaaat cccagaaatc ttgtaggaaa tcagaacccc aacctgaccg tgcagcagag
781 gaatcgcggg agacacagct attggtgtca ggcagggcaa gagcaaagtc tagtggaagc
841 acccctgtta ctgcagcctc ttcacccaaa gtaggaaaga tctggactga gagatggcgc
901 ggtggaatgg tgcctgtcca gacttccaca gagacagcta aaatgaagac ccctgtgcgg
961 cattcacagc aacttaagga tgaagactct cgtgttactg gcagacgaca ttctgtgaat
1021 ctggatgaag gtggaagtgc ccaggcagtc cataaaacag tcactcctgg gaaactggcg
1081 actagaaacc aaactccggt ggaggctggg gatgttggca gccccgctga tacaccagaa
1141 cattcctctt ccccccagag aagtattcct gcaaaggtag aggctccatc tgcagagaca
1201 caaaatcggc tctctttaac tcagcgcctt gttccaggtg aaaagaaaac tcccaagggt
1261 tccttcagca agcctgagaa actggccaca gccgccgaac agacttgctc tggcctacct
1321 ggtcttagtt ccgttgatat cagcaacttt ggtgattcca ttaacaagag tgagggaatg
1381 cctatgaaga gaagacgtgt atcctttggt ggacatctaa gacctgaatt atttgatgaa
1441 aacttgcctc ctaatacacc actgaaaaga ggagaaacgc caaccaagag gaagtctctt
1501 ggcactcaca gcccagctgt cctcaagaca atcatcaagg aacggcccca gtctccaggg
1561 aaacaagagt ctcctgggat aacgccaccg aggacaaatg atcaaagacg cagatcaggc

1621 aggacttcca gtggaagcaa tttcttatgt gagacagaca ttcccaagaa agcaggcagg
1681 aagagcggta acctgcctgc gaagagagca tccatcagcc ggagtcagca tggcattcta
1741 cagatgattt gctccaaaag gcgaagtgga gcttctgaag ccaacttgat tgttgcaaaa
1801 tcatgggctg atgttgtaaa acttggcgtg aaacaaacac aaacgaaagt tgcgaaacat
1861 gtccctccaa agcagacgag caagagacaa agaagaccca gcactccaaa gaaacccaca
1921 agcaatcttc acaatcaatt tactacaggc catgcaaact ctccctgtac cattgtagta
1981 ggtagagcgc agattgaaaa agtaagtgtg cctgcccgac cctacaaaat gctgaataac
2041 ttgatgctaa accgaaaagt ggacttcagt gaagatctgt caggactaac tgaaatgttc
2101 aagactccag tgaaggagaa gcagcagcag atgagtgata caggctccgt actttccaat
2161 tcagcgaatt tgtctgaaag acaattgcaa gtaactaatt caggagacat acctgagccc
2221 atcaccacag agattttggg agaaaaagtg ctatccagta ctcggaatgc agcaaagcag
2281 cagtctgata gatattctgc aagtcctacc ttaagacggc ggagcatcaa acatgaaaac
2341 acagtgcaaa ctcctaagaa tgtccataac attactgacc ttgagaagaa gactccggtc
2401 tctgagacag agcccctgaa gactgcatcg agtgtgagca agttaagaag atctagagag
2461 ctcagacata cccttgtgga aactatgaat gaaaaaacag aagcagtcct tgctgagaac
2521 accacagcaa gacatttaag ggggacattt cgagaacaaa aagtagatca acaggtgcag
2581 gacaatgaaa acgctcctca aagatgcaag gaaagtggtg aattaagtga aggttcagaa
2641 aagacatcag ctaggagatc aagtgccagg aagcagaagc cgacaaaaga cttactagga
2701 agtcagatgg tcacccaaac agcagactat gctgaggaac tacttagtca aggacaagga
2761 accatacaaa acctagagga atccatgcac atgcaaaaca catcaataag tgaggatcaa
2821 ggaattacag aaaagaaagt gaacataata gtatatgcaa ccaaagagaa gcactcgcca
2881 aagacccctg gcaaaaaggc acaacctcta gaagggccag ctggtctcaa ggaacacttt
2941 gaaacaccaa accccaaaga taaacctata acggaagaca gaactagagt cctttgcaaa
3001 tcaccacaag tcacaacaga gaatatcaca acaaacacaa agccacagac tagcacatct
3061 gggaagaaag tagacatgaa ggaagaaagc tctgccttga caaaacgtat acatatgcca
3121 ggggaatcca ggcataatcc caaaatttta aaacttgagt gtgaggatat caaagctttg
3181 aagcaatctg aaaatgaaat gctgacctca acagtaaatg gaagcaagag gactttagga
3241 aaatctaaaa aaaaggctca gcccctggaa gacctgactt gtttccagga actctttata
3301 tcaccagttc ctactaacat aatcaaaaaa attcccagca aatctccaca cacacaacca
3361 gtcagaaccc cagcgagcac aaagagactc tccaagacag gtctcagtaa agtggatgtg
3421 agacaagaac cttcaacact tgggaaaaga acgaagtcac caggcagagc cccaggcaca
3481 ccagcaccag tgcaggaaga aaatgactgc acagcctaca tggaaactcc aaagcagaaa
3541 ctggagtcta tagaaaattt aacagggctt aggaaacagt ccagaacacc taaagacatc
3601 actggtttcc aggatagttt ccaaatacca gatcatgcta atggcccatt agtggttgtc
3661 aaaaccaaaa aaatgttctt taattctcca caaccagaaa gtgccataac ccgaaagagc
3721 agagagagac agtctagggc aagtataagt aaaatagatg ttaaagaaga acttttagaa
3781 tcagaggaac acctacaatt aggagaaggt gtagacacat ttcaggtatc caccaacaaa
3841 gtcattagat catctaggaa acctgcaaag cgtaaactgg attcaacagc tggtatgcct
3901 aacagcaaga ggatgcgctg ttcttcaaag gataacacac catgcctaga agacctgaat
3961 ggcttccaag agctcttcca aatgccaggc tatgctaatg actctttgac cactggaatc
4021 tcaacaatgc ttgctagatc accacaatta ggaccagtta gaacccaaat caacaaaaag
4081 agtctgccca agatcatctt gagaaaaatg gatgtgacag aagaaatttc aggtctctgg
4141 aagcagtcac tgggcagagt ccacaccaca caagagcagg aggataatgc aatcaaagca
4201 attatggaga ttccaaagga aacactgcag actgcagcag atggaactag gcttaccaga

4261 cagccacaaa cacctaagga aaaagttcaa ccgctggaag atcacagtgt cttccaagaa
4321 ctcttccaaa catcacgcta ctgttctgat ccattaattg gtaacaaaca aacaagaatg
4381 tccttgagat ctccacaacc aggatttgtt agaactccac gaacctcaaa gagactggct
4441 aagacaagtg ttgggaatat tgctgtgaga gaaaagatct ctccagtgag tctgccacag
4501 tgtgctacag gggaggttgt acacataccc atagggccag aagatgacac agagaacaaa
4561 ggtgtgaagg aatccacacc tcagacactg gactcatcag caagtcgaac tgtcagcaag
4621 aggcagcaag gggcacatga ggaaaggcct cagttctcag gagacttatt tcatccccaa
4681 gagctctttc aaacaccagc cagtggcaaa gacccagtaa ctgttgatga aactacaaaa
4741 atagctctgc agtctccaca accaggacat atcataaacc cagcaagcat gaagagacag
4801 tccaacatga gtctcaggaa agacatgaga gaattttcca tacttgaaaa acaaacacag
4861 tcacgaggca gagacgcagg cacaccagca ccaatgcagg aagaaaatgg caccacagcc
4921 attatggaaa caccaaagca gaaactggat ttcataggaa attcaacagg acataagagg
4981 aggcctcgga cacccaaaaa cagggctcag cccctagaag acctggatgg cttccaagaa
5041 ctctttcaaa caccagctgg tgccagtgac cctgtgagtg ttgaagaaag tgcaaagata
5101 tctttggcat cttcacaagc agaaccagtc agaaccccag caagtacaaa gagacgctcc
5161 aagacaggtc tcagtaaagt ggatgtgaga caagaacctt caacacttgg gaaaagaatg
5221 aagtcactag gcagagcccc aggcacacca gcaccagtgc aggaagaaaa tgacagcaca
5281 gccttcatgg aaactccaaa gcagaaactg gatttcacag gaaattcatc aggacataag
5341 aggaggccac agacacctaa gatcagggct cagcccctag aagacctgga tggcttccaa
5401 gaactcttcc aaacaccagc tggtgccaat gactcagtga ctgttgagga aagtgtaaag
5461 atgtctttgg aatcttcaca agcagaacca gtcaaaaccc cggcaagcac aaagagactc
5521 tccaagacag gtctcagtaa ggtggatgtg agagaagacc cttcaatact tgagaaaaaa
5581 acaaagtcac caggcacacc agcaccagtg caggaagaaa atgactgcac agccttcatg
5641 gaaactccaa agcagaaact ggatttcaca ggaaattcat caggacataa gaggaggcca
5701 cggacaccta agatcagagc tcagccccta gaagacctgg atggcttcca agaactcttc
5761 caaacaccag ctggtgctag tgactcagtg actgttgagg aaagtgcaaa gatgtctttg
5821 gaatcttcac aagcaaaacc agtcaaaacc ccggcaagca caaagagact ctccaagaca
5881 ggtctcagta aggtggatgt gagagaagac ccttcaacac ttgggaaaaa aacaaagtca
5941 ccaggcagag ccccaggcac accagcacca gtgcaggaag aaaatgacag cacagccttc
6001 atggaaactc caaagcagaa actggatttt gcagagaatt catcagggag taagagaagg
6061 tcacgaacat ctaagaacag gtctcagccc ctagaagacc tggatggctt ccaagaactc
6121 ttccaaacac cagctggtgc cagtaaccct gtgagtgttg aagaaagtgc aaagatatct
6181 ttggaatctt cacaagcaga accagtcaga acccgggcaa gcacaaagag actttccaag
6241 acaggtctca ataagatgga tgtgagagaa gggcactctc cgctcagtaa gtcaagctgt
6301 gcatcacaga aagtcatgca aaccctcaca cttggagaag atcatggcag agagaccaaa
6361 gatgggaagg tattgttagc tcagaaattg gaaccagcaa tatatgttac tcgtggcaag
6421 aggcagcaaa ggtcatgtaa gaaaaggtcc cagtccccag aagacctctc tggtgttcag
6481 gaggtcttcc aaacatcagg ccataacaag gattcagtga cagtggacaa tcttgcaaaa
6541 ctgcccagct cgtctccacc actagagcca acagacactt cagtaacctc acggagacag
6601 gccagaactg gtctgaggaa agttcacgtg aaaaatgaac tttcaggagg cataatgcat
6661 ccacaaatat caggggaaat tgtggactta cctagagaac cagaaggtga aggcaaagtc
6721 attaaaacaa ggaagcaatc tgtaaaacgg aaattggaca cagaagtcaa tgtgcctcgc
6781 agtaagaggc aaagaattac aagagcagaa aagaccctag aggatctgcc tggcttccaa
6841 gagctctgcc aagctccaag cttggtaatg gactcagtta ttgttgagaa aaccccaaag

6901 atgcccgaca aatctccaga acctgtggat acaacttcag agacacaggc aagaagaaga
6961 ctcaggagac tggttgttac tgaagagccc ataccacaaa gaaagactac aagagttgta
7021 aggcaaacca gaaacacaca gaaagagccc ataagtgaca atcaaggtat ggaagagttt
7081 aaggaatctt cagtacagaa acaagaccca agtgtaagtt taactggcag gaggaaccaa
7141 ccaaggacag ttaaggagaa aacccaaccc ttagaagaac tcaccagttt ccaagaggaa
7201 actgccaaaa gaatatcttc caaatctcca caaccggaag agaaggaaac cttagcaggt
7261 ttaaagaggc agctcagaat acaactaatc aacgatggtg taaaagaaga gcccacagca
7321 cagagaaagc aaccatccag ggaaaccagg aacacactca aagagcctgt aggtgacagt
7381 ataaatgttg aagaggttaa gaagtctaca aagcagaaaa ttgatccagt agcaagtgtg
7441 cctgtcagca agaggccacg gagggtaccc aaggaaaagg cacaggccct agaattggct
7501 ggtctcaaag gaccaatcca aaccctaggc cacactgatg aatcagcaag tgataaagga
7561 cccacacaga tgccctgtaa ttctctacaa ccagagcaag ttgacagctt ccaaagctca
7621 ccaaggcgac ccaggacaag acgtgggaaa gtagaggcag atgaagagcc ttcagcagta
7681 agaaagacag tatcaacatc aaggcaaact atgcgatccc gcaaggtccc tgaaattggt
7741 aacaatggta cccaagtttc aaaggcctcc ataaagcaga cattagatac agtagccaaa
7801 gtaactggca gcaggaggca gctaaggaca cataaaggat ggggttcaac cctcttgaag
7861 ttgttaggtg actccaaaga aataacccaa atatcagatc actctgagaa actagcacat
7921 gacaccagta tccttaagag cactcaacag caaaagccag actcagtaaa acctctgaga
7981 acatgcagaa gagtgctgag ggcctctaaa gaggtcccca aggaagtgtt ggtggacacc
8041 agagaccatg caacattaca aagcaaaagc aacccttttgc tgtccccgaa gaggaagtct
8101 gcaagagatg gaagcattgt gagaaccagg gctttgcgct ctttagcacc aaagcaggaa
8161 gcaacagatg agaagcctgt acctgagaaa aaaagggctg cttccagcaa gaggtatgta
8221 tcacctgagc ctgtgaagat gaaacacctg aaaatcgtgt caaacaaact tgaatctgtg
8281 gaagagcagg ttagcactgt tatgaaaaca gaagaaatgg aagccaaaag agaaaatcct
8341 gtcactccag atcagaactc taggtaccga aagaaaacca atgtaaaaca gccaaggccc
8401 aagtttgatg catctgcaga gaatgtcggg ataaagaaaa acgagaagac tatgaagact
8461 gcctcccagg agacagagct gcagaatcca gatgatggag ccaagaaatc tacatctcgg
8521 ggccaagtca gtgggaaaag aacatgcttg aggtctagag gaacgactga gatgccccag
8581 ccttgtgaag cagaagagaa aacaagcaaa ccagctgcag aaatcttgat aaagcctcag
8641 gaagagaaag gagtctctgg agagtctgat gttaggtgtt tgaggtccag aaaaactaga
8701 gtcgctttgg acagtgaacc taagccaagg gtaactcgtg gaaccaagaa agatgcaaaa
8761 actctgaagg aggatgaaga cattgtatgc accaagaagt taagaacaag aagttaa

//

8. MKI67 (*Mus musculus*) Protein sequence

1 massahlvti krsgddgahf plslssclfg rsiecdiriq lpvvsqrhcp ivvqeqeail
61 ynfsstnptq vngvtidepv rlrhgdiiti idrsfryedg nhedgskpte fpgkslgkep
121 srrasrdsfc adpdgegqdt kaskmtasrr sfvyakglsa dspasdgskn svsqdssghv
181 eqhtgrnive ptsggsllrs pglqgavtgn rsllptqsls nsnekespfe klyqsmkeel
241 dvksqkscrk sepqpdraae esretqllvs graraksgs tpvtaasspk vgkiwterwr
301 ggmvpvqtst etakmktpvr hsqqlkdeds rvtgrrhsvn ldeggsaqav hktvtpgkla
361 trnqtpveag dvgspadtpe hssspqrsip akveapsaet qnrlsltqrl vpgekktpkg
421 sfskpeklat aaeqtcsglp glssvdisnf gdsinksegm pmkrrrvsfg ghlrpelfde
481 nlppntplkr getptkrksl gthspavlkt iikerpqspg kqespgitpp rtndqrrrsg
541 rtssgsnflc etdipkkagr ksgnlpakra sisrsqhgil qmicskrrsg aseanlivak

601 swadvvklgv kqtqtkvakh vppkqtskrq rrpstpkkpt snlhnqfttg hanspctivv
661 graqiekvsv parpykmlnn lmlnrkvdfs edlsgltemf ktpvkekqqq msdtgsvlsn
721 sanlserqlq vtnsgdipep itteilgekv lsstrnaakq qsdrysaspt lrrrsikhen
781 tvqtpknvhn itdlekktpv seteplktas svsklrrsre lrhtlvetmn ekteavlaen
841 ttarhlrgtf reqkvdqqvq dnenapqrck esgelsegse ktsarrssar kqkptkdllg
901 sqmvtqtady aeellsqgqg tiqnleesmh mqntsisedq gitekkvnii vyatkekhsp
961 ktpgkkaqpl egpaglkehf etpnpkdkpi tedrtrvlck spqvttenit tntkpqtsts
1021 gkkvdmkees saltkrihmp gesrhnpkil klecedikal kqsenemlts tvngskrtlg
1081 kskkkaqple dltcfqelfi spvptniikk ipsksphtqp vrtpastkrl sktglskvdv
1141 rqepstlgkr tkspgrapgt papvqeendc taymetpkqk lesienltgl rkqsrtpkdi
1201 tgfqdsfqip dhangplvvv ktkkmffnsp qpesaitrks rerqsrasis kidvkeelle
1261 seehlqlgeg vdtfqvstnk virssrkpak rkldstagmp nskrmrcssk dntpcledln
1321 gfqelfqmpg yandslttgi stmlarspql gpvrtqinkk slpkiilrkm dvteeisglw
1381 kqslgrvhtt qeqednaika imeipketlq taadgtrltr qpqtpkekvq pledhsvfqe
1441 lfqtsrycsd plignkqtrm slrspqpgfv rtprtskrla ktsvgniavr ekispvslpq
1501 catgevvhip igpeddtenk gvkestpqtl dssasrtvsk rqqgaheerp qfsgdlfhpq
1561 elfqtpasgk dpvtvdettk ialqspqpgh iinpasmkrq snmslrkdmr efsilekqtq
1621 srgrdagtpa pmqeengtta imetpkqkld fignstghkr rprtpknraq pledldgfqe
1681 lfqtpagasd pvsveesaki slassqaepv rtpastkrrs ktglskvdvr qepstlgkrm
1741 kslgrapgtp apvqeendst afmetpkqkl dftgnssghk rrpqtpkira qpledldgfq
1801 elfqtpagan dsvtveesvk mslessqaep vktpastkrl sktglskvdv redpsilekk
1861 tkspgtpapv qeendctafm etpkqkldft gnssghkrrp rtpkiraqpl edldgfqelf
1921 qtpagasdsv tveesakms1 essqakpvkt pastkrlskt glskvdvred pstlgkktks
1981 pgrapgtpap vqeendstaf metpkqkldf aenssgskrr srtsknrsqp ledldgfqel
2041 fqtpagasnp vsveesakis lessqaepvr trastkrlsk tglnkmdvre ghsplskssc
2101 asqkvmqtlt lgedhgretk dgkvllaqkl epaiyvtrgk rqqrsckkrs qspedlsgvq
2161 evfqtsghnk dsvtvdnlak lpssspplep tdtsvtsrrq artglrkvhv knelsggimh
2221 pqisgeivdl prepegegkv iktrkqsvkr kldtevnvpr skrqritrae ktledlpgfq
2281 elcqapslvm dsvivektpk mpdkspepvd ttsetqarrr lrrlvvteep ipqrkttrvv
2341 rqtrntqkep isdnqgmeef kessvqkqdp svsltgrrnq prtvkektqp leeltsfqee
2401 takrissksp qpeeketlag lkrqlriqli ndgvkeepta qrkqpsretr ntlkepvgds
2461 inveevkkst kqkidpvasv pvskrprrvp kekaqalela glkgpiqtlg htdesasdkg
2521 ptqmpcnslq peqvdsfqss prrprtrrgk veadeepsav rktvstsrqt mrsrkvpeig
2581 nngtqvskas ikqtldtvak vtgsrrqlrt hkgwgstllk llgdskeitq isdhseklah
2641 dtsilkstqq qkpdsvkplr tcrrvlrask evpkevlvdt rdhatlqsks npllspkrks
2701 ardgsivrtr alrslapkqe atdekpvpek kraasskryv spepvkmkhl kivsnklesv
2761 eeqvstvmkt eemeakrenp vtpdqnsryr kktnvkqprp kfdasaenvg ikknektmkt
2821 asqetelqnp ddgakkstsr gqvsgkrtcl rsrgttempq pceaeektsk paaeilikpq
2881 eekgvsgesd vrclrsrktr valdsepkpr vtrgtkkdak tlkededivc tkklrtrs

//

9. CCNB1 (*Homo sapiens*) cDNA sequence

1 atggcgctcc gagtcaccag gaactcgaaa attaatgctg aaaataaggc gaagatcaac
61 atggcaggcg caaagcgcgt tcctacggcc cctgctgcaa cctccaagcc cggactgagg
121 ccaagaacag ctcttgggga cattggtaac aaagtcagtg aacaactgca ggccaaaatg

```
 181 cctatgaaga aggaagcaaa accttcagct actggaaaag tcattgataa aaaactacca
 241 aaacctcttg aaaaggtacc tatgctggtg ccagtgccag tgtctgagcc agtgccagag
 301 ccagaacctg agccagaacc tgagcctgtt aaagaagaaa aactttcgcc tgagcctatt
 361 ttggttgata ctgcctctcc aagcccaatg gaaacatctg gatgtgcccc tgcagaagaa
 421 gacctgtgtc aggctttctc tgatgtaatt cttgcagtaa atgatgtgga tgcagaagat
 481 ggagctgatc caaacctttg tagtgaatat gtgaaagata tttatgctta tctgagacaa
 541 cttgaggaag agcaagcagt cagaccaaaa tacctactgg gtcgggaagt cactggaaac
 601 atgagagcca tcctaattga ctggctagta caggttcaaa tgaaattcag gttgttgcag
 661 gagaccatgt acatgactgt ctccattatt gatcggttca tgcagaataa ttgtgtgccc
 721 aagaagatgc tgcagctggt tggtgtcact gccatgttta ttgcaagcaa atatgaagaa
 781 atgtaccctc cagaaattgg tgactttgct tttgtgactg acaacactta tactaagcac
 841 caaatcagac agatggaaat gaagattcta agagctttaa actttggtct gggtcggcct
 901 ctacctttgc acttccttcg gagagcatct aagattggag aggttgatgt cgagcaacat
 961 actttggcca aatacctgat ggaactaact atgttggact atgacatggt gcactttcct
1021 ccttctcaaa ttgcagcagg agctttttgc ttagcactga aaattctgga taatggtgaa
1081 tggacaccaa ctctacaaca ttacctgtca tatactgaag aatctcttct tccagttatg
1141 cagcacctgg ctaagaatgt agtcatggta aatcaaggac ttacaaagca catgactgtc
1201 aagaacaagt atgccacatc gaagcatgct aagatcagca ctctaccaca gctgaattct
1261 gcactagttc aagatttagc caaggctgtg gcaaaggtgt aa
```

//

10. CCNB1 (*Homo sapiens*) Protein sequence

```
   1 malrvtrnsk inaenkakin magakrvpta paatskpglr prtalgdign kvseqlqakm
  61 pmkkeakpsa tgkvidkklp kplekvpmlv pvpvsepvpe pepepepepv keeklspepi
 121 lvdtaspspm etsgcapaee dlcqafsdvi lavndvdaed gadpnlcsey vkdiyaylrq
 181 leeeqavrpk yllgrevtgn mrailidwlv qvqmkfrllq etmymtvsii drfmqnncvp
 241 kkmlqlvgvt amfiaskyee myppeigdfa fvtdntytkh qirqmemkil ralnfglgrp
 301 lplhflrras kigevdveqh tlakylmelt mldydmvhfp psqiaagafc lalkildnge
 361 wtptlqhyls yteesllpvm qhlaknvvmv nqgltkhmtv knkyatskha kistlpqlns
 421 alvqdlakav akv
```

//

11. CCNB1 (*Mus musculus*) cDNA sequence

```
   1 atggccctca gggtcactag gaacacgaaa attaacgcag aaaataaggc caaggtcagt
  61 atggcaggcg caatgcgtgt gcctgtgaca gttactgctg cttccaagcc cgggctgaga
 121 ccgagaactg ctcttggaga cattggtaat aaagtcagcg aagagctaca ggcaagagtg
 181 cctctgaaaa gggaagcaaa aacgctaggt actggaaaag gtactgttaa agccctacca
 241 aaacctgtag agaaggtgcc tgtgtgtgaa ccagaggtgg aacttgctga gcctgagcct
 301 gaacctgaac ttgaacatgt tagagaagag aagctttctc ctgaacctat tttggttgat
 361 aatccctctc caagcccgat ggaaacatct ggatgtgcgc ctgcagaaga gtatctgtgt
 421 caggctttct ctgatgtaat ccttgcagtg agtgacgtag acgcagatga tggggctgac
 481 ccaaacctct gtagtgaata tgtgaaagat atctatgctt atctccgaca actggaggaa
 541 gagcagtcag ttagaccaaa atacctacag ggtcgtgaag tgactggaaa catgagagct
 601 atcctcattg actggctaat acaggttcag atgaaattta ggctgcttca ggagaccatg
 661 tacatgactg tgtccattat tgatcggttc atgcagaaca gttgtgtgcc caagaagatg
 721 ctacagctgg tcggtgtaac ggccatgttt attgcaagca aatatgagga gatgtaccct
```

781 ccagaaatag gtgacttcgc ctttgtgact aacaacacgt acactaagca ccagatcaga
841 cagatggaga tgaagattct cagagttctg aacttcagcc tgggtcgccc tctgcctctg
901 cacttcctcc gtagagcatc taaagtcgga gaggttgacg tcgagcagca cactttggcc
961 aaatacctca tggagctctc catgctggac tgcgacatgg tgcattttgc tccttctcaa
1021 attgcagctg gggctttctg cttagcgctg aaaattcttg acaacggtga atggacacca
1081 actctgcagc actacctatc ctacagtgaa gactccctgc ttcctgttat gcagcacctg
1141 gctaagaatg tagtcatggt gaactgtggc ctcacaaagc acatgactgt caagaacaag
1201 tatgcagcat ctaagcatgc taagatcagc acgctggcac agctgaactg tacactagtt
1261 cagaatttgt ctaaggccgt gacaaaggca taa

//

12. CCNB1 (*Mus musculus*) Protein sequence

1 malrvtrntk inaenkakvs magamrvpvt vtaaskpglr prtalgdign kvseelqarv
61 plkreaktlg tgkgtvkalp kpvekvpvce pevelaepep epelehvree klspepilvd
121 npspspmets gcapaeeylc qafsdvilav sdvdaddgad pnlcseyvkd iyaylrqlee
181 eqsvrpkylq grevtgnmra ilidwliqvq mkfrllqetm ymtvsiidrf mqnscvpkkm
241 lqlvgvtamf iaskyeemyp peigdfafvt nntytkhqir qmemkilrvl nfslgrplpl
301 hflrraskvg evdveqhtla kylmelsmld cdmvhfapsq iaagafclal kildngewtp
361 tlqhylsyse dsllpvmqhl aknvvmvncg ltkhmtvknk yaaskhakis tlaqlnctlv
421 qnlskavtka

//

13. BUB1 (*Homo sapiens*) cDNA sequence

1 atggacaccc cggaaaatgt ccttcagatg cttgaagccc acatgcagag ctacaagggc
61 aatgaccctc ttggtgaatg ggaaagatac atacagtggg tagaagagaa ttttcctgag
121 aataaagaat acttgataac tttactagaa catttaatga aggaattttt agataagaag
181 aaataccaca atgacccaag attcatcagt tattgtttaa aatttgctga gtacaacagt
241 gacctccatc aatttttga gtttctgtac aaccatggga ttggaaccct gtcatcccct
301 ctgtacattg cctgggcggg gcatctggaa gcccaaggag agctgcagca tgccagtgct
361 gtccttcaga gaggaattca aaaccaggct gaacccagag agttcctgca acaacaatac
421 aggttatttc agacacgcct cactgaaacc catttgccag ctcaagctag aacctcagaa
481 cctctgcata atgttcaggt tttaaatcaa atgataacat caaaatcaaa tccaggaaat
541 aacatggcct gcatttctaa gaatcagggt tcagagcttt ctggagtgat atcttcagct
601 tgtgataaag agtcaaatat ggaacgaaga gtgatcacga tttctaaatc agaatattct
661 gtgcactcat ctttggcatc caaagttgat gttgagcagg ttgttatgta ttgcaaggag
721 aagcttattc gtggggaatc agaatttttcc tttgaagaat tgagagccca gaaatacaat
781 caacggagaa agcatgagca atgggtaaat gaagacagac attatatgaa aaggaaagaa
841 gcaaatgctt ttgaagaaca gctattaaaa cagaaaatgg atgaacttca taagaagttg
901 catcaggtgg tggagacatc ccatgaggat ctgcccgctt cccaggaaag gtccgaggtt
961 aatccagcac gtatggggcc aagtgtaggc tcccagcagg aactgagagc gccatgtctt
1021 ccagtaacct atcagcagac accagtgaac atggaaaaga acccaagaga ggcacctcct
1081 gttgttcctc ctttggcaaa tgctatttct gcagctttgg tgtccccagc caccagccag
1141 agcattgctc ctcctgttcc tttgaaagcc cagacagtaa cagactccat gtttgcagtg
1201 gccagcaaag atgctggatg tgtgaataag agtactcatg aattcaagcc acagagtgga
1261 gcagagatca aagaagggtg tgaaacacat aaggttgcca acacaagttc ttttcacaca
1321 actccaaaca catcactggg aatggttcag gcaacgccat ccaaagtgca gccatcaccc

```
1381 accgtgcaca caaaagaagc attaggtttc atcatgaata tgtttcaggc tcctacactt
1441 cctgatattt ctgatgacaa agatgaatgg caatctctag atcaaaatga agatgcattt
1501 gaagcccagt ttcaaaaaaa tgtaaggtca tctggggctt ggggagtcaa taagatcatc
1561 tcttctttgt catctgcttt tcatgtgttt gaagatggaa acaaagaaaa ttatggatta
1621 ccacagccta aaaataaacc cacaggagcc aggacctttg gagaacgctc tgtcagcaga
1681 cttccttcaa aaccaaagga ggaagtgcct catgctgaag agtttttgga tgactcaact
1741 gtatggggta ttcgctgcaa caaaaccctg gcacccagtc ctaagagccc aggagacttc
1801 acatctgctg cacaacttgc gtctacacca ttccacaagc ttccagtgga gtcagtgcac
1861 attttagaag ataaagaaaa tgtggtagca aaacagtgta cccaggcgac tttggattct
1921 tgtgaggaaa acatggtggt gccttcaagg gatggaaaat tcagtccaat tcaagagaaa
1981 agcccaaaac aggccttgtc gtctcacatg tattcagcat ccttacttcg tctgagccag
2041 cctgctgcag gtggggtact tacctgtgag gcagagttgg gcgttgaggc ttgcagactc
2101 acagacactg acgctgccat tgcagaagat ccaccagatg ctattgctgg gctccaagca
2161 gaatggatgc agatgagttc acttgggact gttgatgctc caaacttcat tgttgggaac
2221 ccatgggatg ataagctgat tttcaaactt ttatctgggc tttctaaacc agtgagttcc
2281 tatccaaata cttttgaatg gcaatgtaaa cttccagcca tcaagcccaa gactgaattt
2341 caattgggtt ctaagctggt ctatgtccat caccttcttg gagaaggagc ctttgcccag
2401 gtgtacgaag ctacccaggg agatctgaat gatgctaaaa ataaacagaa atttgtttta
2461 aaggtccaaa agcctgccaa cccctgggaa ttctacattg ggacccagtt gatggaaaga
2521 ctaaagccat ctatgcagca catgtttatg aagttctatt ctgcccactt attccagaat
2581 ggcagtgtat tagtaggaga gctctacagc tatggaacat tattaaatgc cattaacctc
2641 tataaaaata cccctgaaaa agtgatgcct caaggtcttg tcatctcttt tgctatgaga
2701 atgctttaca tgattgagca agtgcatgac tgtgaaatca ttcatggaga cattaaacca
2761 gacaatttca tacttggaaa cggatttttg gaacaggatg atgaagatga tttatctgct
2821 ggcttggcac tgattgacct gggtcagagt atagatatga aacttttttcc aaaaggaact
2881 atattcacag caaagtgtga aacatctggt tttcagtgtg ttgagatgct cagcaacaaa
2941 ccatggaact accagatcga ttactttggg gttgctgcaa cagtatattg catgctcttt
3001 ggcacttaca tgaaagtgaa aaatgaagga ggagagtgta agcctgaagg tctttttaga
3061 aggcttcctc atttggatat gtggaatgaa ttttttcatg ttatgttgaa tattccagat
3121 tgtcatcatc ttccatcttt ggatttgtta aggcaaaagc tgaagaaagt atttcaacaa
3181 cactatacta acaagattag ggccctacgt aataggctaa ttgtactgct cttagaatgt
3241 aagcgttcac gaaaataa
```

//

14. BUB1 (*Homo sapiens*) Protein sequence

```
   1 mdtpenvlqm leahmqsykg ndplgewery iqwveenfpe nkeylitlle hlmkefldkk
  61 kyhndprfis yclkfaeyns dlhqffefly nhgigtlssp lyiawaghle aqgelqhasa
 121 vlqrgiqnqa epreflqqqy rlfqtrltet hlpaqartse plhnvqvlnq mitsksnpgn
 181 nmacisknqg selsgvissa cdkesnmerr vitiskseys vhsslaskvd veqvvmycke
 241 klirgesefs feelraqkyn qrrkheqwvn edrhymkrke anafeeqllk qkmdelhkkl
 301 hqvvetshed lpasqersev nparmgpsvg sqqelrapcl pvtyqqtpvn meknpreapp
 361 vvpplanais aalvspatsq siappvplka qtvtdsmfav askdagcvnk sthefkpqsg
 421 aeikegceth kvantssfht tpntslgmvq atpskvqpsp tvhtkealgf imnmfqaptl
 481 pdisddkdew qsldqnedaf eaqfqknvrs sgawgvnkii sslssafhvf edgnkenygl
 541 pqpknkptga rtfgersvsr lpskpkeevp haeeflddst vwgircnktl apspkspgdf
```

```
 601 tsaaqlastp fhklpvesvh iledkenvva kqctqatlds ceenmvvpsr dgkfspiqek
 661 spkqalsshm ysasllrlsq paaggvltce aelgveacrl tdtdaaiaed ppdaiaglqa
 721 ewmqmsslgt vdapnfivgn pwddklifkl lsglskpvss ypntfewqck lpaikpktef
 781 qlgsklvyvh hllgegafaq vyeatqgdln daknkqkfvl kvqkpanpwe fyigtqlmer
 841 lkpsmqhmfm kfysahlfqn gsvlvgelys ygtllnainl ykntpekvmp qglvisfamr
 901 mlymieqvhd ceiihgdikp dnfilgngfl eqddeddlsa glalidlgqs idmklfpkgt
 961 iftakcetsg fqcvemlsnk pwnyqidyfg vaatvycmlf gtymkvkneg geckpeglfr
1021 rlphldmwne ffhvmlnipd chhlpsldll rqklkkvfqq hytnkiralr nrlivlllec
1081 krsrk
```

//

15. BUB1 (*Mus musculus*) cDNA sequence

```
   1 atggacaacc tagaaaatgt ctttcgcatg tttgaagccc atatgcaaag ctacacgggt
  61 aatgacccac ttggagaatg ggaaagcttt ataaagtggg tagaagagaa ttttcctgac
 121 aataaagaat acttgatgac attattagaa catttaatga aggaattttt acataagaag
 181 aactaccaca atgattcaag attcatcaat tattgcttaa aatttgctga gtacaacagc
 241 gaccgtcatc agttttttga gtttctgtac aaccagggaa ttggaaccaa gtcatcatat
 301 atatacatgt cctgggcagg gcatctggaa gcccagggag agctgcagca tgccagtgct
 361 atttttcaga caggaattca caatgaggct gaacctaaag aactactaca gcaacaatac
 421 aggctattcc aagcacgcct tactggaatc catttgccag ctcaagctac aacctcagaa
 481 cctttgcata gtgcacagat tttaaaccaa gttatgatga caaactcaag tccagaaaaa
 541 aactcagcct gtgttcctaa gagtcagggt tcagaatgtt ctggtgtggc atcttccact
 601 tgtgatgaaa agtctaatat ggaacaaagg gtgatcatga tttccaagtc agaatgctct
 661 gtcagctcat ctgtggcacc caagcctgag gctcagcaag ttatgtactg caaggaaaag
 721 cttattcgtg gagattcaga attttctttt gaagaactga gagcccagaa atataatcaa
 781 aggaagaagc atgagcagtg ggttagtgaa gacagaaatt atatgaaaag gaaagaagca
 841 aatgcttttg aagagcaatt attaaaacag aaaatggatg aacttcacaa gaaattgcat
 901 caagtggtgg aattgtcaca caaggacctt cctgcttctg agaacaggcc tgatgttagt
 961 ctagtatgtg ttggacaaaa tacttgctcc cagcaggaat tgaggggtcc aagtctttca
1021 tccatcagtc atcagacctc agagagttca ggagagaaac cacaggaaga accttctgtt
1081 cctcttatgg taaatgctgt taacagcact ttgctgttcc cagctgccaa cctgccagct
1141 cttcctgttc ctgtaagtgg ccagtcattg acagactcca gatgtgtgaa tcaaagtgtt
1201 catgaattca tgccacagtg tggaccagaa acaaaagaag tgtgtgaaac aaataaagtt
1261 gccagcatta atgattttca tacaactcca aacacatcat tgggaatggt tcaaggaaca
1321 ccatgcaaag tgcagccatc accaactgtc cacaccaagg aagcattagg tttcatcatg
1381 gacatgtttc aggctccaac acttcctgac atttctgatg ataaagatga atggccatct
1441 ctggaccaaa atgaagatgc atttgaagcc cagtttcaaa aaaatgcagt atcttcggga
1501 gattggggag ttaaaaaaat tatgactttg tcatctgctt ttcctatttt tgaagatgga
1561 aacaaagaaa attatggctt accacagcct aaaaataagc ccttaggagc taggaccttt
1621 ggagaacgat ctctcagtaa atattcctcg agatcaaatg aaatgcctca cactgatgag
1681 tttatggatg attcaacagt atgtggtatt cgctgcaaca aaactctagc tcccagtcct
1741 aaaagtatag gagactttac atctgctgcc caactttcgt ctacaccatt ccacaaattt
1801 ccagcagatt tagtacagat tccagaagat aaagaaaatg tggtagccac acagtataca
1861 catatggctt tggattcttg taaagaaaac atagtggacc tctcaaaagg cagaaagctt
1921 gggccaattc aagagaaaat ttcagcatct ttaccctgtc ctagtcagcc tgccacaggt
```

1981 ggtttgttca cccaggaagc agtgttcggc cttgaggctt ttaaatgcac aggcattgac
2041 catgcgacag tggaagacct atccgatgcc aatgctgggc tccaagttga atgcgtgcag
2101 acacttggaa atgtcaatgc tccaagcttt actgttgaga acccatggga tgatgaattg
2161 attcttaaac ttctctctgg actttctaag ccagttactt cctattcaaa tacttttgag
2221 tggcagagta aacttccagc catcaagacc aagacagaat atcaattggg ttctttgctg
2281 gtctatgtga atcaccttct tggagaagga gcctttgctc aagtctttga agctattcat
2341 ggagatgtga gaaatgccaa aagtgaacag aaatgcattt tgaaggtgca gagacctgcc
2401 aactcctggg aattctacat tgggatgcag ctgatggaaa gactaaagcc agaagtacat
2461 cacatgttca tcaagtttta ttctgctcat ttattcaaga acggcagcat attagtaggg
2521 gaactctaca gctatgggac gttactaaat gtcattaacc tctataaaaa tacctctgaa
2581 aaagtgatgc cccaggctct tgtcctcact ttcgctatca gaatgcttta catggttgaa
2641 caagtccaca gctgcgaaat cattcatgga gacattaagc cagataactt catactagga
2701 cacagatttt tggaacaggc tgatgaagac ttagctaccg gcttggcatt gattgacctg
2761 ggtcagagta tagatatgaa acttttccct aaaggaactg tatttacagg aaaatgtgaa
2821 acatctggtt ttcagtgtcc tgagatgctc agtaacaagc catggaacta ccagattgat
2881 tactttggag ttgctgcaac aatatactgt atgctctttg gctcttacat gaaagtaaaa
2941 aatgaaggag gagtctggaa acctgaaggt ctttttagaa ggcttcctca tttggatatg
3001 tgggaggaat tttttcacat catgttgaat ataccggatt gtcataatct tccatctttg
3061 gattttctga gacagaatat gaagaaatta cttgaacaac agtattccaa caagattaag
3121 accttgcgta ataggctaat tgtgatgctt tcagaatata agcgttcaag aaaataa

//

16. BUB1 (*Mus musculus*) Protein sequence

1 mdnlenvfrm feahmqsytg ndplgewesf ikwveenfpd nkeylmtlle hlmkeflhkk
61 nyhndsrfin yclkfaeyns drhqffefly nqgigtkssy iymswaghle aqgelqhasa
121 ifqtgihnea epkellqqqy rlfqarltgi hlpaqattse plhsaqilnq vmmtnsspek
181 nsacvpksqg secsgvasst cdeksnmeqr vimisksecs vsssvapkpe aqqvmyckek
241 lirgdsefsf eelraqkynq rkkheqwvse drnymkrkea nafeeqllkq kmdelhkklh
301 qvvelshkdl pasenrpdvs lvcvgqntcs qqelrgpsls sishqtsess gekpqeepsv
361 plmvnavnst llfpaanlpa lpvpvsgqsl tdsrcvnqsv hefmpqcgpe tkevcetnkv
421 asindfhttp ntslgmvqgt pckvqpsptv htkealgfim dmfqaptlpd isddkdewps
481 ldqnedafea qfqknavssg dwgvkkimtl ssafpifedg nkenyglpqp knkplgartf
541 gerslskyss rsnemphtde fmddstvcgi rcnktlapsp ksigdftsaa qlsstpfhkf
601 padlvqiped kenvvatqyt hmaldscken ivdlskgrkl gpiqekisas lpcpsqpatg
661 glftqeavfg leafkctgid hatvedlsda naglqvecvq tlgnvnapsf tvenpwddel
721 ilkllsglsk pvtsysntfe wqsklpaikt kteyqlgsll vyvnhllgeg afaqvfeaih
781 gdvrnakseq kcilkvqrpa nswefyigmq lmerlkpevh hmfikfysah lfkngsilvg
841 elysygtlln vinlykntse kvmpqalvlt fairmlymve qvhsceiihg dikpdnfilg
901 hrfleqaded latglalidl gqsidmklfp kgtvftgkce tsgfqcpeml snkpwnyqid
961 yfgvaatiyc mlfgsymkvk neggvwkpeg lfrrlphldm weeffhimln ipdchnlpsl
1021 dflrqnmkkl leqqysnkik tlrnrlivml seykrsrk

//

17. KNTC2 (*Homo sapiens*) cDNA sequence

1 atgaagcgca gttcagtttc cagcggtggt gctggccgcc tctccatgca ggagttaaga
61 tcccaggatg taaataaaca aggcctctat acccctcaaa ccaaagagaa accaaccttt

```
 121 ggaaagttga gtataaacaa accgacatct gaaagaaaag tctcgctatt tggcaaaaga
 181 actagtggac atggatcccg gaatagtcaa cttggtatat tttccagttc tgagaaaatc
 241 aaggacccga gaccacttaa tgacaaagca ttcattcagc agtgtattcg acaactctgt
 301 gagtttctta cagaaaatgg ttatgcacat aatgtgtcca tgaaatctct acaagctccc
 361 tctgttaaag acttcctgaa gatcttcaca tttctttatg gcttcctgtg cccctcatac
 421 gaacttcctg acacaaagtt tgaagaagag gttccaagaa tctttaaaga ccttgggtat
 481 ccttttgcac tatccaaaag ctccatgtac acagtggggg ctcctcatac atggcctcac
 541 attgtggcag ccttagtttg gctaatagac tgcatcaaga tacatactgc catgaaagaa
 601 agctcacctt tatttgatga tgggcagcct tggggagaag aaactgaaga tggaattatg
 661 cataataagt tgtttttgga ctacaccata aaatgctatg agagttttat gagtggtgcc
 721 gacagctttg atgagatgaa tgcagagctg cagtcaaaac tgaaggattt atttaatgtg
 781 gatgctttta agctggaatc attagaagca aaaaacagag cattgaatga acagattgca
 841 agattggaac aagaaagaga aaaagaaccg aatcgtctag agtcgttgag aaaactgaag
 901 gcttccttac aaggagatgt tcaaaagtat caggcataca tgagcaattt ggagtctcat
 961 tcagccattc ttgaccagaa attaaatggt ctcaatgagg aaattgctag agtagaacta
1021 gaatgtgaaa caataaaaca ggagaacact cgactacaga atatcattga caaccagaag
1081 tactcagttg cagacattga gcgaataaat catgaaagaa atgaattgca gcagactatt
1141 aataaattaa ccaaggacct ggaagctgaa caacagaagt tgtggaatga ggagttaaaa
1201 tatgccagag gcaaagaagc gattgaaaca caattagcag agtatcacaa attggctaga
1261 aaattaaaac ttattcctaa aggtgctgag aattccaaag gttatgactt tgaaattaag
1321 tttaatcccg aggctggtgc caactgcctt gtcaaataca gggctcaagt ttatgtacct
1381 cttaaggaac tcctgaatga aactgaagaa gaaattaata aagccctaaa taaaaaaatg
1441 ggtttggagg atactttaga acaattgaat gcaatgataa cagaaagcaa gagaagtgtg
1501 agaactctga aagaagaagt tcaaaagctg gatgatcttt accaacaaaa aattaaggaa
1561 gcagaggaag aggatgaaaa atgtgccagt gagcttgagt ccttggagaa acacaagcac
1621 ctgctagaaa gtactgttaa ccaggggctc agtgaagcta tgaatgaatt agatgctgtt
1681 cagcgggaat accaactagt tgtgcaaacc acgactgaag aaagacgaaa agtgggaaat
1741 aacttgcaac gtctgttaga gatggttgct acacatgttg ggtctgtaga gaaacatctt
1801 gaggagcaga ttgctaaagt tgatagagaa tatgaagaat gcatgtcaga agatctctcg
1861 gaaaatatta aagagattag agataagtat gagaagaaag ctactctaat taagtcttct
1921 gaagaatga
```

//

18. KNTC2 (*Homo sapiens*) Protein sequence

```
   1 mkrssvssgg agrlsmqelr sqdvnkqgly tpqtkekptf gklsinkpts erkvslfgkr
  61 tsghgsrnsq lgifssseki kdprplndka fiqqcirqlc efltengyah nvsmkslqap
 121 svkdflkift flygflcpsy elpdtkfeee vprifkdlgy pfalskssmy tvgaphtwph
 181 ivaalvwlid cikihtamke ssplfddgqp wgeetedgim hnklfldyti kcyesfmsga
 241 dsfdemnael qsklkdlfnv dafkleslea knralneqia rleqerekep nrleslrklk
 301 aslqgdvqky qaymsnlesh saildqklng lneeiarvel ecetikqent rlqniidnqk
 361 ysvadierin hernelqqti nkltkdleae qqklwneelk yargkeaiet qlaeyhklar
 421 klklipkgae nskgydfeik fnpeagancl vkyraqvyvp lkellnetee einkalnkkm
 481 gledtleqln amiteskrsv rtlkeevqkl ddlyqqkike aeeedekcas eleslekhkh
 541 llestvnqgl seamneldav qreyqlvvqt tteerrkvgn nlqrllemva thvgsvekhl
 601 eeqiakvdre yeecmsedls enikeirdky ekkatlikss ee
```

//

19. KNTC2 (*Mus musculus*) cDNA sequence

```
   1 atgaagcgca gttcagtttc cacctgtggt gctggccgcc tctctatgca ggagttaagg
  61 accctggacc tcaataagcc aggcctttat acccctcaaa ccaaagaaag atcaaccttt
 121 ggaaagctga gtacacacaa accgacatcg gaaagaaaag tctcaatatt tgggaaaagg
 181 actagcggac atggatccag gaatagtcaa cttggtatat tttccagttc tgaaaaaatc
 241 aaggacccaa gaccacttaa tgacaaagca ttcattcagc agtgtattcg acaactctat
 301 gagtttctta cagaaaacgg ttatgtgtat agtgtatcca tgaagtctct gcaagctcca
 361 tccactaaag agttcctaaa gatcttcgcc tttctttatg gctttctgtg cccgtcgtat
 421 gaacttcctg gtacaaaatg tgaagaagag gtcccaagaa tttttaaagc acttgggtat
 481 cccttcacac tgtccaagag ctccatgtat acagtgggag cccctcacac gtggcctcac
 541 atcgtggctg ccttggtgtg gctcatagac tgcatcaaga ttgatactgc catgaaagaa
 601 agctcacctt tatttgatga tgggcagctc tggggagaag agactgaaga tggaattaaa
 661 cacaataagt tgtttttgga gtacaccaaa aagtgctatg agaagttcat gaccggggcc
 721 gacagctttg aagaagagga tgctgagctg caggcgaagc tgaaggactt gtacaaggta
 781 gatgcatcta agctggagtc actcgaagca gaaaacaaag aactaaatga acagattgca
 841 agactggagg aggaaagaga aagagaaccg aaccgtctga tgtcattgaa gaaactgaaa
 901 gcgtccttac aagcagatgt tcaaaactat aaagcataca tgagcaactt ggagtctcat
 961 ttagccgttc tgaaacagaa atcgaatagt cttgatgaag aaattggtag agtagaacaa
1021 gaatgtgaaa ctgttaaaca ggaaaacact cgactacaga gtatcgttga taaccagaag
1081 tattcagtcg ctgacattga gagaataaat catgagaaaa atgaattgca gcagactatt
1141 aataaattaa ccaaagacct ggaagccgaa cagcaacaga tgtggaatga agaattaaaa
1201 tacgcaagag gcaaagaggc gattgaagcg cagctagcgg agtaccacaa gttggctaga
1261 aaattaaagc ttatccccaa aggtgctgag aattccaaag gttacgactt tgaaattaag
1321 tttaatcctg aggcgggtgc caactgcctt gtcaaataca ggactcaagt gtatgcaccg
1381 ctcaaagagc tcttgaatga aagcgaagaa gaaattaaca aagctctgaa taaaaagagg
1441 catctggagg atactttaga acaactgaac accatgaaaa cggaaagcaa gaacactgtg
1501 aggatgctga aggaggagat tcagaaactg gatgaccttc accagcaggc agtgaaggaa
1561 gctgaggaaa aagacaagaa gagtgccagt gagcttgagt ccctggagaa acacaagcac
1621 ctgctggaga gcggggtgaa cgatggcctc agcgaggcca tggatgagtt ggacgctgtc
1681 cagcgggaat accagctaac tgtgaagacc acaactgaag aaagaagaaa ggtggaaaac
1741 aacttacaac gtcttttgga gatggtcgcc acacacgtag ggtctttgga gaaacatctt
1801 gaagaggaga atgctaaagc cgacagagag tacgaagaat tcatgtctga agatctcctg
1861 gaaaacatca gggagatggc agagaagtat aagagaaatg ctgcccaact taaggctccc
1921 gacaaatga
```

//

20. KNTC2 (Mus musculus) Protein sequence

```
   1 mkrssvstcg agrlsmqelr tldlnkpgly tpqtkerstf gklsthkpts erkvsifgkr
  61 tsghgsrnsq lgifssseki kdprplndka fiqqcirqly efltengyvy svsmkslqap
 121 stkeflkifa flygflcpsy elpgtkceee vprifkalgy pftlskssmy tvgaphtwph
 181 ivaalvwlid cikidtamke ssplfddgql wgeetedgik hnklfleytk kcyekfmtga
 241 dsfeeedael qaklkdlykv daskleslea enkelneqia rleeererep nrlmslkklk
 301 aslqadvqny kaymsnlesh lavlkqksns ldeeigrveq ecetvkqent rlqsivdnqk
 361 ysvadierin heknelqqti nkltkdleae qqqmwneelk yargkeaiea qlaeyhklar
```

421 klklipkgae nskgydfeik fnpeagancl vkyrtqvyap lkellnesee einkalnkkr
481 hledtleqln tmktesknty rmlkeeiqkl ddlhqqavke aeekdkksas eleslekhkh
541 llesgvndgl seamdeldav qreyqltvkt tteerrkven nlqrllemva thvgslekhl
601 eeenakadre yeefmsedll eniremaeky krnaaqlkap dk

//

21. USP22 (*Homo sapiens*) cDNA sequence

1 atggcgccgg gttggccctc actatcagcg ggctcccgac aggaggcgcc ccagcttgcg
61 gccgggggca gcgcctacca ggcagttggc aggcagttcc agccccgggc cacggcactg
121 cagggcccga gccaggttac ggcaaccggg ggccctgcaa acagctcccg attagggggt
181 gcttttgggt gggaaagcgc aggccctgga tggaggcccg acctgcggcg ctccagctca
241 tcgccccgcc tctttgcggc agagctcagg ccggcgcaaa ccggttcagt gctaggactg
301 acgtctcgcg ccgcccaacc gcagcacgcc cccgcctccc cagtcctctg gaagagacag
361 ggaacgtcta gccgccaggg tcccgggagg cggctctgta ccagacggac tatactgaga
421 gcctatgaca atagccgaag agcgcagcgc aggcggtccg cagcagccgc agctcggggg
481 cggtgcctgc cttgcagcct cccctcggcg atcgcgcagc cccatctttg tccggcctcc
541 gcgctttgtt ctcggcgccc gggccttggc cagcctggcc agccgccgag cagcccccac
601 gccgcgctgg cgtcgtcctc gcctccctcg ccgccgcccc ccgcgcgcgg ccgggccttg
661 ccccccatgg tgtcccggcc agagcccgag ggcgaggcca tggacgccga gctggcggta
721 gcgccgccgg gctgctcgca cctgggcagc ttcaaggtgg acaactggaa gcagaacctg
781 cgggccatct accagtgctt cgtgtggagc ggcacggctg aggcccgcaa gcgcaaggcc
841 aagtcctgta tctgccatgt ctgtggcgtc cacctcaaca ggctgcattc ctgcctctac
901 tgtgtcttct tcggctgttt cacaaagaag catattcacg agcatgcgaa ggcgaagcgg
961 cacaacctgg ccattgatct gatgtacgga ggcatctact gttttctgtg ccaggactac
1021 atctatgaca aagacatgga aataatcgcc aaggaggagc agcgaaaagc ttggaaaatg
1081 caaggcgttg gagagaagtt ttcaacttgg gaaccaacca aacgggagct tgaactgctg
1141 aagcacaacc cgaaaaggag aaagatcacc tcgaactgca ccataggtct gcgtgggctg
1201 atcaaccttg ggaacacatg cttcatgaac tgcatcgtgc aggccctgac ccacacgcca
1261 cttctgcggg acttcttcct gtctgacagg caccgctgtg agatgcagag ccccagctcc
1321 tgtctggtct gtgagatgtc ctcactgttt caggagtttt actctggaca ccggtccgct
1381 cacatcccgt ataagttgct gcacctggtg tggacccacg cgaggcacct agcaggctac
1441 gagcagcagg acgcccacga gttcctcatc gcggccctgg acgtgctcca ccgacactgc
1501 aaaggtgatg acaatgggaa gaaggccaac aaccccaacc actgcaactg catcatagac
1561 cagatcttca caggcgggtt gcagtcagac gtcacctgcc aagtctgcca tggagtctcc
1621 accaccatcg acccgttctg ggacatcagc ttggatctcc ccggctcttc caccccattc
1681 tggcccctga gcccagggag cgagggcaac gtggtaaacg gggaaagcca cgtgtcggga
1741 accaccacgc tcacggactg cctgcgacga ttcaccagac cagagcactt gggcagcagc
1801 gccaagatca agtgcagcgg ttgccatagc taccaggagt ccacaaagca gctcactatg
1861 aagaaactgc ccatcgtagc ctgttttcat ctcaaacgat ttgaacactc agccaagctg
1921 cggcggaaga tcaccacgta tgtgtccttc cccctggagc tggacatgac ccctttcatg
1981 gcctccagca aagagagcag gatgaatgga cagtaccagc agcccacgga cagtctcaac
2041 aatgacaaca agtattccct gtttgctgtt gttaaccatc aagggacctt ggagagtggc
2101 cactacacca gctttatccg gcagcacaaa gaccagtggt tcaagtgtga cgatgccatc
2161 atcaccaagg ccagcatcaa ggacgtcctg gacagcgaag ggtacttgct gttctatcac
2221 aaacagttcc tggaatacga gtag

//

22. USP22 (*Homo sapiens*) Protein sequence

```
   1 mapgwpslsa gsrqeapqla aggsayqavg rqfqpratal qgpsqvtatg gpanssrlgg
  61 afgwesagpg wrpdlrrsss sprlfaaelr paqtgsvlgl tsraaqpqha paspvlwkrq
 121 gtssrqgpgr rlctrrtilr aydnsrraqr rrsaaaaarg rclpcslpsa iaqphlcpas
 181 alcsrrpglg qpgqppssph aalassspps ppppargral ppmvsrpepe geamdaelav
 241 appgcshlgs fkvdnwkqnl raiyqcfvws gtaearkrka kscichvcgv hlnrlhscly
 301 cvffgcftkk hihehakakr hnlaidlmyg giycflcqdy iydkdmeiia keeqrkawkm
 361 qgvgekfstw eptkrelell khnpkrrkit snctiglrgl inlgntcfmn civqalthtp
 421 llrdfflsdr hrcemqspss clvcemsslf qefysghrsp hipykllhlv wtharhlagy
 481 eqqdahefli aaldvlhrhc kgddngkkan npnhcnciid qiftgglqsd vtcqvchgvs
 541 ttidpfwdis ldlpgsstpf wplspgsegn vvngeshvsg tttltdclrr ftrpehlgss
 601 akikcsgchs yqestkqltm kklpivacfh lkrfehsakl rrkittyvsf pleldmtpfm
 661 asskesrmng qyqqptdsln ndnkyslfav vnhqgtlesg hytsfirqhk dqwfkcddai
 721 itkasikdvl dsegyllfyh kqfleye
```

//

23. USP22 (*Mus musculus*) cDNA sequence

```
   1 atggtggcca ggccggagcc tgaggtcgag gccatggacg ctgagctggc ggtaccgccg
  61 cctggctgct cgcacctggg cagcttcaag gtggacaact ggaagcaaaa cctgcgggcc
 121 atctaccagt gcttcgtgtg gagcggaact gccgaggctc gcaagcgcaa ggcaaagtcc
 181 tgtgtctgcc atgtctgcgg catccacctg aaccggctgc actcttgcct ctactgtgtc
 241 ttctttggct gtttcacgaa gaagcacatc catgaccatg ccaagtcaaa gcgacacaac
 301 ctggccatcg acctgatgta cggaggtatt tactgcttct tgtgtcagga ctacatctat
 361 gacaaagaca tagaaatcat tgccaaagag gagcagcgca aggcttggaa gatgcaaggt
 421 gttggagaga agttttcaac ttgggaacca actaaacggg agctggaact gctgaagcat
 481 aacccaaaga ggcggaagat cacctccaat tgtaccatag gtctgcgtgg actgatcaac
 541 ctggggaaca cgtgtttcat ggactgcatc gtgcaggcgc tgacccacac tccgctcctg
 601 agagacttct ttctgtcgga taggcaccgc tgtgagatgc agagccccag ctcctgcttg
 661 gtctgtgaga tgtcctctct cttccaggag ttttactcag ggcaccgctc cccacacatt
 721 ccatacaagc tgctgcacct ggtgtggacg cacgcccggc acctggcggg ttatgagcag
 781 caggacgcac atgagttcct cattgcagcc ctggacgtcc tccaccggca ctgcaaaggt
 841 gatgacaatg ggaagaaagc caacaatcct aaccactgca attgcatcat tgaccagatc
 901 tttacgggtg ggctccagtc tgatgttaca tgccaagtct gccacggggt ctccaccacc
 961 atagacccct tctgggacat cagtttagac cttcccggtt cttctacccc attctggccc
1021 ttgagcccag ggagcgaggg cagtgtggtt aatggggaga gccatgcatc cgggaccacc
1081 actctcacag actgcctgcg aagatttacc agaccagagc acttaggaag cagtgccaag
1141 atcaagtgta gcggttgcca tagctaccaa gagtccacaa agcagctcac catgaagaag
1201 ctgcccattg tggcctgttt ccatctcaaa cgatttgaac actcagccaa acttcggcgg
1261 aagatcacca catatgtgtc ttttcccctg gaactggaca tgacgccctt catggcctcc
1321 agcaaagaga gcaggatgaa tgggcaatac cagcagcccc tggacagtct caacaatgac
1381 aacaaatact ccctgtttgc tgtcgttaac catcaaggga ccttggagag tggccactac
1441 accagcttca tccggcagca caaagaccag tggttcaagt gtgatgacgc cattatcacc
1501 aaggccagca tcaaagatgt actggacagt gaagggtacc tactcttcta tcacaaacag
1561 ttcctggaat acgagtag
```

//

24. USP22 (*Mus musculus*) Protein sequence

```
   1 mvarpepeve amdaelavpp pgcshlgsfk vdnwkqnlra iyqcfvwsgt aearkrkaks
  61 cvchvcgihl nrlhsclycv ffgcftkkhi hdhakskrhn laidlmyggi ycflcqdyiy
 121 dkdieiiake eqrkawkmqg vgekfstwep tkrelellkh npkrrkitsn ctiglrglin
 181 lgntcfmdci vqalthtpll rdfflsdrhr cemqspsscl vcemsslfqe fysghrsphi
 241 pykllhlvwt harhlagyeq qdahefliaa ldvlhrhckg ddngkkannp nhcnciidqi
 301 ftgglqsdvt cqvchgvstt idpfwdisld lpgsstpfwp lspgsegsvv ngeshasgtt
 361 tltdclrrft rpehlgssak ikcsgchsyq estkqltmkk lpivacfhlk rfehsaklrr
 421 kittyvsfpl eldmtpfmas skesrmngqy qqpldslnnd nkyslfavvn hqgtlesghy
 481 tsfirqhkdq wfkcddaiit kasikdvlds egyllfyhkq fleye
```

//

25. HCFC1 (*Homo sapiens*) cDNA sequence

```
   1 atggtggagt atgggaaata cagcaatgac ctctacgaac tccaggcgag ccggtgggag
  61 tggaagagac tcaaagcaaa gacgcccaaa aacgggcccc ctccgtgtcc tcgactcggg
 121 cacagcttct cccttgtggg caacaaatgc tacctgtttg ggggtctggc caatgatagc
 181 gaggacccaa agaacaacat tccaaggtac ctgaatgact tatatatcct ggaattacgg
 241 ccaggctctg gagtggtagc ctgggacatt cccatcactt acggggtcct accaccaccc
 301 cgggagtcac atactgccgt ggtctacacc gaaaaagaca ataagaagtc caagctggtg
 361 atctacggcg ggatgagtgg ctgcaggctg ggggacctgt ggaccctaga tattgacacc
 421 ctgacgtgga ataagcccag tctcagcggg gtggcgcctc ttcctcgcag tctccactcg
 481 gcaaccacca tcggaaataa aatgtacgtg tttggtggct gggtgcctct cgtcatggat
 541 gacgtcaaag tggccacaca cgagaaggag tggaagtgta ccaacacgct ggcttgtctc
 601 aacctggata ccatggcctg ggagaccatc ctgatggata cactggagga caacatcccc
 661 cgtgctcggg ctggccactg cgcagtcgcc atcaacaccc gcctgtacat ttggagtggg
 721 cgtgacggct accgcaaggc ctggaacaac caggtctgct gcaaggacct ctggtaccta
 781 gagacagaaa agccaccacc cccagcccga gtacaactgg tacgcgccaa caccaactcc
 841 ctggaggtga gctgggggggc agtggcaaca gccgacagct accttctcca gctccagaaa
 901 tatgacattc ctgccacggc tgctactgcc acctcccctа cacccaatcc ggtcccatct
 961 gtgcctgcca accctcccaa gagccctgcc ccagcagcag ccgcacctgc tgtgcagccg
1021 ctgacccaag taggcatcac gctcctgccc caggctgccc ccgcaccccc gaccaccacc
1081 accatccagg tcttgccaac ggtgcctggc agctccattt ctgtgcccac cgcagccagg
1141 actcaaggtg tccctgctgt tctcaaagtg accggtcctc aggctacaac aggaactcca
1201 ttggtcacca tgcgacctgc cagccaggct gggaaagccc ctgtcaccgt gacctccctt
1261 cccgccggag tgcggatggt tgtgccaaca cagagtgccc agggaacggt gattggcagt
1321 agcccacaga tgagtgggat ggccgcactg gccgctgcgg ccgctgccac ccagaagatc
1381 ccccttcct cgcgacccac ggtgctgagt gtcccagcgg gtaccaccat cgtgaagacc
1441 atggctgtga cacctggcac taccaccctc ccagccactg tgaaggtggc ctcctcgcca
1501 gtcatggtga gcgtgagcaa ccctgccact cgcatgctga agactgcagc cgcccaggtg
1561 gggacatcgg tttcctccgc caccaacacg tctacccgcc ctatcatcac agtgcacaag
1621 tcaggcactg tgacagtggc ccagcaagcc caggtggtga ccacagttgt gggcggggtc
1681 accaagacca tcaccctggt gaagagcccc atctctgtcc caggaggcag tgctctgatt
1741 tccaatctgg gcaaagtgat gtcggtggtc cagaccaaac cagttcagac ttcagcagtc
1801 acaggccagg cgtccacggg tcctgtgact cagatcatcc agaccaaagg gcccctgcca
```

1861 gcgggaacaa tcctgaagct ggtgacctca gcagatggca agcccaccac catcatcact
1921 accacgcagg ccagtggggc ggggaccaag cccaccatcc tgggcatcag cagcgtctcc
1981 cccagtacca ccaagcccgg cacgaccacc atcatcaaaa ccatccccat gtcggccatc
2041 atcacccagg cgggcgccac gggtgtgacc agcagtcctg gcatcaagtc acccatcacc
2101 atcatcacca ccaaggtgat gacttcagga actggagcac ctgcgaaaat catcactgct
2161 gtccccaaaa ttgccactgg ccacgggcag cagggagtga cccaggtggt gcttaagggg
2221 gccccgggac agccaggcac catcctccgc actgtgccca tgggggggtgt tcgcctggtc
2281 acacccgtca ccgtctccgc cgtcaagcca gccgtcacca cgttggttgt gaaaggcacc
2341 acaggtgtca cgaccctagg cacagtgaca ggcaccgtct ccaccagcct tgccggggcg
2401 gggggccaca gcactagtgc ttccctggcc acgcccatca ccaccttggg caccattgcc
2461 accctctcaa gccaggtgat caaccccact gccatcactg tgtcggccgc acagaccacg
2521 ctgacagcgg caggcgggct cacaaccccg accatcacca tgcagcccgt gtcccagccc
2581 acccaggtaa ctctgatcac ggcacctagt ggggtggagg cccagcctgt gcatgacctc
2641 cctgtgtcca ttctggcctc cccgactaca gaacagccca ccgccacagt taccatcgcc
2701 gactcaggcc agggtgatgt gcagcctggc actgtcacct tggtgtgctc caacccaccc
2761 tgtgagaccc acgagactgg caccaccaac acggccacca ctactgttgt ggctaacctt
2821 gggggacacc cccagcccac ccaagtgcag ttcgtctgtg acagacagga ggcagctgct
2881 tctcttgtga cctcgactgt gggccagcag aatggtagcg tggtccgagt ctgttcgaac
2941 ccgccctgcg agacccacga gacgggcacc accaacaccg ccaccaccgc cacctccaac
3001 atggccgggc agcatggctg ctcaaaccca ccctgcgaga cccacgagac gggcaccacc
3061 aacactgcca ctacagccat gtcgagcgtc ggcgccaacc accagcgaga tgcccgtcgg
3121 gcctgtgcag ctggcacccc tgccgtgatc cggatcagtg tggccactgg ggcgctggag
3181 gcagcccagg gctctaagcc ccagtgccaa acccgccaga ccagcgcgac cagcaccacc
3241 atgactgtga tggccaccgg ggccccgtgc tcggccggcc cactccttgg gccgagcatg
3301 gcacgggagc ccgggggccg cagccctgct tttgtgcagt tggcccctct gagcagcaaa
3361 gtcaggctga gcagcccaag cattaaggac cttcctgcgg ggcgccacag ccatgcggtc
3421 agcaccgctg ccatgacccg ttccagcgtg ggtgctgggg agccccgcat ggcacctgtg
3481 tgcgagagcc tccagggtgg ctcgcccagc accacagtga ctgtgacagc cctggaggca
3541 ctgctgtgcc cctcggccac cgtgacccaa gtctgctcca acccaccatg tgagacccac
3601 gagacaggca ccaccaacac cgccactacc tcgaatgcag gcagcgccca gagggtgtgc
3661 tccaacccgc catgcgagac ccacgagacg ggcaccaccc acacggccac caccgctact
3721 tcaaacgggg gcacgggcca gcccgagggt gggcagcagc cccctgctgg tcgcccctgt
3781 gagacacacc agaccacttc cactggcacc accatgtcgg tcagcgtggg tgccctgctt
3841 cccgacgcca cttcttccca caggaccgtg gagtctggcc tagaggtggc ggcggcaccc
3901 agcgtcaccc cccaggctgg caccgcgctg ctggctcctt tcccaacaca gagggtgtgc
3961 tccaaccccc cctgtgagac ccacgagacg ggcaccactc acacggccac cactgtcact
4021 tccaacatga gttcaaacca agacccccca cctgctgcca gcgatcaggg agaggtggag
4081 agcacccagg gcgacagcgt gaacatcacc agctccagtg ccatcacgac aaccgtgtcc
4141 tccacactga cgcgggctgt gaccaccgtg acgcagtcca caccggtccc gggcccctct
4201 gtgccgcccc cagaggaact ccaggtgtcg ccaggtcctc gccagcagct gccaccacgg
4261 cagcttctgc agtcggcttc cacagccctg atgggggagt ccgccgaggt cctgtcagcc
4321 tcccagaccc ctgagctccc ggccgccgtg gatctgagca gcacagggga gccatcttcg
4381 ggccaggagt ctgccggctc tgcggtggtg gccactgtgg tggtccagcc acccccaccc
4441 acacagtccg aagtagacca gttatcactt ccccaagagc taatggccga ggcccaagct

4501 ggcaccacca ccctcatggt aacggggctc acccccgagg agctggcagt gacggctgct
4561 gcagaagcag ctgcccaggc cgcagccacg gaggaagccc aggccctggc catccaggcg
4621 gtgctccagg ccgcgcagca ggccgtcatg ggcaccggcg agcccatgga cacctccgag
4681 gcagcagcaa ccgtgactca ggcggagctg gggcacctgt cggccgaggg tcaggagggc
4741 caggccacca ccatacccat tgtgctgaca cagcaggagc tggctgccct ggtgcagcag
4801 cagcagctgc aggaggccca ggcccagcag cagcatcacc acctccccac tgaggccctg
4861 gcccctgccg acagtctcaa cgacccagcc attgagagca attgcctcaa tgagctggcc
4921 ggcacggtcc ccagcactgt ggcgctgctg ccctcaacgg ccactgagag cctggctcca
4981 tccaacacat ttgtggcccc ccagccggtt gtggtggcca gcccagccaa gctgcaggct
5041 gcagctaccc tgaccgaagt ggccaatggc atcgagtccc tgggtgtgaa gccagacctg
5101 ccgcccccac ccagcaaagc ccccatgaag aaggaaaacc agtggtttga tgtgggagtc
5161 attaagggca ccaatgtaat ggtgacacac tatttcctgc caccagatga tgctgtccca
5221 tcagacgatg atttgggcac cgtccctgac tataaccagc tgaagaagca ggagctgcag
5281 ccaggcacag cctataagtt tcgtgttgcc ggaatcaatg cctgtgcgcg ggggcccttc
5341 agcgaaatct cagcctttaa gacgtgcctg cctggtttcc caggggcccc ttgtgccatt
5401 aaaatcagca aaagtccgga tggtgctcac ctcacctggg agccaccctc tgtgacctcc
5461 ggcaagatta tcgagtactc cgtgtacctg gccatccaga gctcacaggc tgggggcgag
5521 ctcaagagct ccaccccggc ccagctggcc ttcatgcggg tgtactgcgg gcccagcccc
5581 tcctgcctgg tgcagtcctc cagcctttcc aacgcccaca tcgactacac caccaagccc
5641 gccatcatct tccgcatcgc cgcccgcaat gagaagggct atggcccggc cacacaagtg
5701 aggtggctgc aggaaaccag taaagacagc tctggcacca agccagccaa caagcggccc
5761 atgtcctctc cagaaatgaa atctgctcca aagaaatcta aggccgatgg tcagtga

//

26. HCFC1 (*Homo sapiens*) Protein sequence

1 mveygkysnd lyelqasrwe wkrlkaktpk ngpppcprlg hsfslvgnkc ylfgglands
61 edpknnipry lndlyilelr pgsgvvawdi pitygvlppp reshtavvyt ekdnkksklv
121 iyggmsgcrl gdlwtldidt ltwnkpslsg vaplprslhs attignkmyv fggwvplvmd
181 dvkvatheke wkctntlacl nldtmaweti lmdtlednip raraghcava intrlyiwsg
241 rdgyrkawnn qvcckdlwyl etekppppar vqlvrantns levswgavat adsyllqlqk
301 ydipataata tsptpnpvps vpanppkspa paaaapavqp ltqvgitllp qaapappttt
361 tiqvlptvpg ssisvptaar tqgvpavlkv tgpqattgtp lvtmrpasqa gkapvtvtsl
421 pagvrmvvpt qsaqgtvigs spqmsgmaal aaaaaatqki ppssrptvls vpagttivkt
481 mavtpgtttl patvkvassp vmvsvsnpat rmlktaaaqv gtsvssatnt strpiitvhk
541 sgtvtvaqqa qvvttvvggv tktitlvksp isvpggsali snlgkvmsvv qtkpvqtsav
601 tgqastgpvt qiiqtkgplp agtilklvts adgkpttiit ttqasgagtk ptilgissvs
661 psttkpgttt iiktipmsai itqagatgvt sspgikspit iittkvmtsg tgapakiita
721 vpkiatghgq qgvtqvvlkg apgqpgtilr tvpmggvrlv tpvtvsavkp avttlvvkgt
781 tgvttlgtvt gtvstslaga gghstsasla tpittlgtia tlssqvinpt aitvsaaqtt
841 ltaagglttp titmqpvsqp tqvtlitaps gveaqpvhdl pvsilasptt eqptatvtia
901 dsgqgdvqpg tvtlvcsnpp cethetgttn tatttvvanl gghpqptqvq fvcdrqeaaa
961 slvtstvgqq ngsvvrvcsn ppcethetgt tntattatsn magqhgcsnp pcethetgtt
1021 ntattamssv ganhqrdarr acaagtpavi risvatgale aaqgskpqcq trqtsatstt
1081 mtvmatgapc sagpllgpsm arepggrspa fvqlaplssk vrlsspsikd lpagrhshav
1141 staamtrssv gageprmapv ceslqggsps ttvtvtalea llcpsatvtq vcsnppceth

1201 etgttntatt snagsaqrvc snppcethet gtthtattat snggtgqpeg gqqppagrpc
1261 ethqttstgt tmsvsvgall pdatsshrtv esglevaaap svtpqagtal lapfptqrvc
1321 snppcethet gtthtattvt snmssnqdpp paasdqgeve stqgdsvnit sssaitttvs
1381 stltravttv tqstpvpgps vpppeelqvs pgprqqlppr qllqsastal mgesaevlsa
1441 sqtpelpaav dlsstgepss gqesagsavv atvvvqpppp tqsevdqlsl pqelmaeaqa
1501 gtttlmvtgl tpeelavtaa aeaaaqaaat eeaqalaiqa vlqaaqqavm gtgepmdtse
1561 aaatvtqael ghlsaegqeg qattipivlt qqelaalvqq qqlqeaqaqq qhhhlpteal
1621 apadslndpa iesnclnela gtvpstvall pstateslap sntfvapqpv vvaspaklqa
1681 aatltevang ieslgvkpdl ppppskapmk kenqwfdvgv ikgtnvmvth yflppddavp
1741 sdddlgtvpd ynqlkkqelq pgtaykfrva ginacargpf seisafktcl pgfpgapcai
1801 kiskspdgah ltweppsvts gkiieysvyl aiqssqagge lksstpaqla fmrvycgpsp
1861 sclvqsssls nahidyttkp aiifriaarn ekgygpatqv rwlqetskds sgtkpankrp
1921 msspemksap kkskadgq

//

27. HCFC1 (*Mus musculus*) cDNA sequence

1 atggcttcgg ctgtgtctcc cgcaaacttg ccagcggtgc ttctgcagcc ccgctggaaa
61 cgggtggtgg gctggtcggg tcccgtgccc cgaccccgcc acggccaccg tgcagtggct
121 atcaaggagc ttatagtggt gtttggcggc ggcaacgagg ggatagtgga cgaactacac
181 gtgtacaaca ctgcaaccaa ccagtggttc atcccagctg tgagagggga tatccctcca
241 gggtgtgcag cctatggctt tgtgtgtgat ggtactcgcc tattggtgtt tggtggaatg
301 gtagagtatg gaaaatacag caacgacctc tatgaactcc aggcaagtcg ttgggaatgg
361 aagagactga aggcaaagac acccaaaaat gggcctcctc catgtcctcg gcttggacat
421 agcttctccc ttgtgggcaa caaatgttac ctgtttgggg gtctggccaa tgatagtgag
481 gaccccaaga acaacattcc gaggtacctg aatgacttat atattctcga actacggcca
541 ggctctggag tggtagcttg ggacatcccc atcacttacg gtgtcctgcc tccaccccgg
601 gagtcacata ctgctgtggt ctacactgaa aaagataaca agaaatccaa gctggtgatc
661 tatggaggga tgagtggctg caggctaggg gacctttgga ccctggacat tgagacactg
721 acatggaata agcccagcct tagtggggtg gcacccccttc ctcgcagcct ccactctgca
781 accaccatag gaaacaaaat gtatgtattt ggtggctggg tgccccttgt catggacgat
841 gtcaaagtgg ccacacacga gaaggagtgg aagtgtacca acacactggc ttgtctcaac
901 ctggatacca tggcctggga aaccatcctg atggatacat tggaggacaa cattcctcga
961 gctcgagcag gccactgtgc tgttgccatc aatactcgtc tgtatatttg gagtggccgt
1021 gatggctacc gcaaggcctg gaacaaccag gtctgctgca aggacttgtg gtatttggag
1081 acagaaaagc caccaccccc agcccgagta caactagtac gagccaacac caactcactg
1141 gaggttagct ggggtgcagt ggcaacagct gacagttacc ttctacaact ccagaaatat
1201 gacattcctg ccacagctgc tacggctacc tcccccactc ccaatccagt cccgtctgtg
1261 cctgccaacc ctcccaagag ccctgcgcca gcagcagctg cacctgctgt acagccactg
1321 acccaagtag gcatcacact tgtgccccag gctgccactg caccccaag cacaaccacc
1381 atccaggtct tgccgacagt gccaggcagc tccatttctg tgcccactgc agccaggact
1441 caaggtgtcc ctgctgttct caaagtgact ggtcctcaag ctacaacagg aacaccactg
1501 gttaccatga gacctgcaag ccaggctgga aaagctcctg tcactgtgac ttccctgcct
1561 gccagtgttc gaatggttgt acccacacag agtgcccagg ggacggtgat cggcagcaac
1621 ccacagatga gtgggatggc cgcattggct gctgctgctg ctgccacaca gaaaatccct
1681 ccatcctcag cacccacggt gctgagtgtc ccagcaggga ccaccatcgt caagacagtg

1741 gctgtgacac ctggcacgac cactcttcca gccactgtga aggtggcctc ctcccctgtc
1801 atggtgagca acccagccac tcgaatgcta aagactgcag ctgcccaagt ggggacatct
1861 gtgtcctctg ctgccaacac atctactcgc cctatcatca cagtacacaa atcaggaact
1921 gtaacagtgg cccagcaagc ccaggtggtg accacggtgg taggtggagt caccaagacc
1981 atcaccctag tgaagagccc catctctgtc ccaggaggca gtgctctgat ttccaatctg
2041 ggaaaagtga tgtcggtggt ccagaccaaa ccagttcaga catcagcagt gacaggccaa
2101 gcatctacag gtcctgtgac tcagatcatc cagaccaaag gacccctgcc agcggggact
2161 atcctgaagc tggtgacatc agcagatggc aagcccacaa ccatcattac caccacacag
2221 gctagtgggg cagggaccaa gcccactatc ctgggcatca gtagtgtttc tcccagcacc
2281 accaaacctg gcacaactac cattattaag accattccta tgtcggccat tatcacccag
2341 gcaggtgcca caggtgttac cagcagtcct ggcattaagt ccccaattac aattatcacc
2401 accaaagtga tgacttcagg aacaggagcg cctgctaaaa tcatcactgc tgtccccaag
2461 attgctactg gccatgggca acaaggagtg acccaggtgg tgctaaaggg ggcccctgga
2521 caaccaggca ccatcctccg tactgtgcct atgggcggcg ttcgcctggt caccctgtc
2581 accgtctctg ctgtcaagcc agctgtcacc acattggttg tgaagggtac cacaggtgtt
2641 acaacgctag gcacagtgac aggcactgtc tccaccagcc tggccggagc tggggcacat
2701 agcaccagtg cttccctggc tacacctatc actaccttgg gcactattgc tacgctctca
2761 agccaggtga tcaaccctac tgctatcaca gtgtcagctg cacagactac actaacagct
2821 gctggtgggc ttaccacacc cacaatcaca atgcagcctg tctcccagcc tacccaggtc
2881 actctgatta cagcacccag tggggttgaa gcacagcctg tacatgacct tcctgtatcc
2941 attttggcct cacctactac agagcagccc acagcaacag tcaccatcgc tgactcaggc
3001 cagggtgatg tgcagcccgg cactgtgaca ctggtgtgtt ccaacccacc ctgtgaaacc
3061 catgaaacag gcaccaccaa cacagctacc accactgttg tggctaacct tggtggacat
3121 cctcaaccta cccaggtgca gtttgtttgt gacagacagg agacagctgc ttcacttgtg
3181 acctcagctg taggacaaca gaatggtaat gtggtccgtg tctgttcaaa ccccccctgt
3241 gagacccatg agacgggcac taccaacact gccacaacag ccacctccaa catggctggg
3301 cagcatggct gctcgaaccc cccctgtgag actcatgaga caggcaccac cagcactgcc
3361 actacagcaa tgtccagcat gggcactggg cagcagcgag acactcgtcg taccactaac
3421 acccccactg tagtgcggat cactgtggct cctggggcat tggagagagt ccagggtacc
3481 gtgaagcctc agtgccaaac ccagcagacc aacatgacca ccaccaccat gactgtgcag
3541 gccactggag ctccatgctc agctggcccc ctgcttaggc caagtgtggc actggagtct
3601 gggagccaca gccctgcctt tgtgcaacta gcccttccaa gtgtcagagt tgggctaagt
3661 ggccccagca gcaaggacat gcccacaggg cgccaaccag agacatatca tacttacaca
3721 actaataccc caaccacaac ccgctctatc atggttgctg gggagcttgg tgcagctcgg
3781 gtggtcccca catctacata tgagagcctc caggcaagct ctcctagcag caccatgact
3841 atgacagccc tagaggcact gctgtgccct tcggctactg tcacccaagt ctgctccaac
3901 ccgccatgtg agacccatga gacgggtacc accaacaccg ccactacctc caatgcgggc
3961 agtgctcagc gagtatgctc caacccgcct tgtgagactc atgagacggg caccacacac
4021 acagctacca ctgccacatc aaatggaggt gcaggccagc ctgagggtgg acaacagcct
4081 gccagtggcc atccctgcga gacacaccag accacttcca ctggcaccac tatgtcagtc
4141 agtgtgggta ccctgattcc tgatgctact tcctctcatg gaaccctgga gtcgggctta
4201 gaggtggtag cagtgcccac tgtcacctcc caggctggtt ccacattgct ggcctctttc
4261 ccaacacaga gggtatgctc caaccctcct tgcgagaccc acgagacagg taccacgcac
4321 acagccacca ctgtcacctc taacatgagc tcaaaccaag accctccacc agctgccagt

4381 gaccaaggag aggtggcaag cacccaaggt gacagcacaa atatcaccag tgccagtgct
4441 atcactacaa gtgtgtcttc tacattgcca cgagcagtga ccactgtgac acagtctaca
4501 ccagtccctg gtccctctgt gccgccccca gaggaactcc aggtctcacc agggcctcgc
4561 cagcagctgc ctccacggca actcctgcag tctgcctcca cacccctgat gggggagtct
4621 actgaggtcc tgtcagcctc ccagacccct gagctccagg ccgccgtgga tctgagcagc
4681 actggggacc catcttcagg ccaggagcct accacctctg ctgtcgtggc cactgtggtg
4741 gtccaaccac ccccacccac acagtctgaa gtagaccagt tatcacttcc ccaagagctg
4801 atggctgaag cccaggcggg caccacaacc cttatggtaa cagggctcac tccagaggag
4861 ctggcagtga ctgctgctgc tgaagcagct gctcaagctg cagccactga agaagcccaa
4921 gccttggcca tccaggctgt gctccaggcc gcacagcagg ctgtcatggg cactggggag
4981 cccatggata catctgaagc agcagcagca gtgacacaag cagaactggg tcacctttca
5041 gctgaaggcc aagagggtca ggctaccacc atacccattg tgttgacaca gcaggagctt
5101 gcagccctgg tgcagcagca gcagcagctc caggaggctc aagctcaagc ccagcaacag
5161 caccatcttc ccactgaggc tctggcccca gctgacagtc tcaatgaccc atccatcgag
5221 agcaactgcc tcaacgagtt agctagtgct gtcccaagca ccgtggcttt gctaccctca
5281 acagctaccg agagcctggc tccatctaac acatttgtgg ctccccagcc tgttgtagct
5341 agtccagcaa agatgcaggc tgcagctacc cttactgaag tggccaatgg cattgagtcc
5401 ctgggtgtga aaccggactt gccacccccа cccagcaaag cccctgtgaa aaaggagaac
5461 cagtggtttg atgtgggggt cattaagggt accagtgtaa tggtgacaca ctattttctg
5521 ccaccagatg atgctgttca gtcagatgat gactcaggca cggtcccaga ctataaccag
5581 ctaaagaagc aggagctaca gccaggcacg gcttacaaat ttcgagttgc tggaatcaat
5641 gcttgtggcc ggggacccct cagtgagatc tcagccttta agacttgtct gcctgggttc
5701 ccaggggctc cttgtgctat taaaatcagc aagagcccag atggtgctca cctcacctgg
5761 gagccaccgt ctgtgacctc cggcaagatc atcgagtact ctgtgtacct ggccatccag
5821 agctcacagg ccagtggtga gccaaagagc tccaccccag cccagctggc cttcatgcga
5881 gtgtactgtg ggcctagccc ttcctgccta gtgcagtcct ccagcctctc caacgcccac
5941 attgactata ctacaaagcc tgccatcatc ttccgcattg ctgcccgcaa tgaaaagggc
6001 tacggccctg ccacacaagt gaggtggttg caagaaacta gtaaagacag ctctggcacc
6061 aagccggcca gcaagcggcc catgtcgtct ccagaaatga aatctgctcc aaagaagtct
6121 aaggctgatg gtcagtga

//

28. HCFC1 (*Mus musculus*) Protein sequence

1 masavspanl pavllqprwk rvvgwsgpvp rprhghrava ikelivvfgg gnegivdelh
61 vyntatnqwf ipavrgdipp gcaaygfvcd gtrllvfggm veygkysndl yelqasrwew
121 krlkaktpkn gpppcprlgh sfslvgnkcy lfgglandse dpknnipryl ndlyilelrp
181 gsgvvawdip itygvlpppr eshtavvyte kdnkksklvi yggmsgcrlg dlwtldietl
241 twnkpslsgv aplprslhsa ttignkmyvf ggwvplvmdd vkvathekew kctntlacln
301 ldtmawetil mdtlednipr araghcavai ntrlyiwsgr dgyrkawnnq vcckdlwyle
361 tekpppparv qlvrantnsl evswgavata dsyllqlqky dipataatat sptpnpvpsv
421 panppkspap aaaapavqpl tqvgitlvpq aatappsttt iqvlptvpgs sisvptaart
481 qgvpavlkvt gpqattgtpl vtmrpasqag kapvtvtslp asvrmvvptq saqgtvigsn
541 pqmsgmaala aaaaatqkip pssaptvlsv pagttivktv avtpgtttlp atvkvasspv
601 mvsnpatrml ktaaaqvgts vssaantstr piitvhksgt vtvaqqaqvv ttvvggvtkt
661 itlvkspisv pggsalisnl gkvmsvvqtk pvqtsavtgq astgpvtqii qtkgplpagt

```
 721 ilklvtsadg kpttiitttq asgagtkpti lgissvspst tkpgtttiik tipmsaiitq
 781 agatgvtssp gikspitiit tkvmtsgtga pakiitavpk iatghgqqgv tqvvlkgapg
 841 qpgtilrtvp mggvrlvtpv tvsavkpavt tlvvkgttgv ttlgtvtgtv stslagagah
 901 stsaslatpi ttlgtiatls sqvinptait vsaaqttlta aggtttptit mqpvsqptqv
 961 tlitapsgve aqpvhdlpvs ilasptteqp tatvtiadsg qgdvqpgtvt lvcsnppcet
1021 hetgttntat ttvvanlggh pqptqvqfvc drqetaaslv tsavgqqngn vvrvcsnppc
1081 ethetgttnt attatsnmag qhgcsnppce thetgttsta ttamssmgtg qqrdtrrttn
1141 tptvvritva pgalervqgt vkpqcqtqqt nmtttttmtvq atgapcsagp llrpsvales
1201 gshspafvql alpsvrvgls gpsskdmptg rqpetyhtyt tntptttrsi mvagelgaar
1261 vvptstyesl qasspsstmt mtaleallcp satvtqvcsn ppcethetgt tntattsnag
1321 saqrvcsnpp cethetgtth tattatsngg agqpeggqqp asghpcethq ttstgttmsv
1381 svgtlipdat sshgtlesgl evvavptvts qagstllasf ptqrvcsnpp cethetgtth
1441 tattvtsnms snqdpppaas dqgevastqg dstnitsasa ittsvsstlp ravttvtqst
1501 pvpgpsvppp eelqvspgpr qqlpprqllq sastplmges tevlsasqtp elqaavdlss
1561 tgdpssgqep ttsavvatvv vqpppptqse vdqlslpqel maeaqagttt lmvtgltpee
1621 lavtaaaeaa aqaaateeaq alaiqavlqa aqqavmgtge pmdtseaaaa vtqaelghls
1681 aegqegqatt ipivltqqel aalvqqqqql qeaqaqaqqq hhlptealap adslndpsie
1741 snclnelasa vpstvallps tateslapsn tfvapqpvva spakmqaaat ltevangies
1801 lgvkpdlppp pskapvkken qwfdvgvikg tsvmvthyfl ppddavqsdd dsgtvpdynq
1861 lkkqelqpgt aykfrvagin acgrgpfsei safktclpgf pgapcaikis kspdgahltw
1921 eppsvtsgki ieysvylaiq ssqasgepks stpaqlafmr vycgpspscl vqssslsnah
1981 idyttkpaii friaarnekg ygpatqvrwl qetskdssgt kpaskrpmss pemksapkks
2041 kadgq
```

//

29. RNF2 (*Homo sapiens*) cDNA sequence

```
   1 atgtctcagg ctgtgcagac aaacggaact caaccattaa gcaaaacatg ggaactcagt
  61 ttatatgagt tacaacgaac acctcaggag gcaataacag atggcttaga aattgtggtt
 121 tcacctcgaa gtctacacag tgaattaatg tgcccaattt gtttggatat gttgaagaac
 181 accatgacta caaaggagtg tttacatcgt ttttgtgcag actgcatcat cacagccctt
 241 agaagtggca acaaagaatg tcctacctgt cggaaaaaac tagtttccaa aagatcacta
 301 aggccagacc caaactttga tgcactcatc agcaaaattt atccaagtcg tgatgagtat
 361 gaagctcatc aagagagagt attagccagg atcaacaagc acaataatca gcaagcactc
 421 agtcacagca ttgaggaagg actgaagata caggccatga acagactgca gcgaggcaag
 481 aaacaacaga ttgaaaatgg tagtggagca gaagataatg gtgacagttc acactgcagt
 541 aatgcatcca cacatagcaa tcaggaagca ggccctagta acaaacggac caaaacatct
 601 gatgattctg ggctagagct tgataataac aatgcagcaa tggcaattga tccagtaatg
 661 gatggtgcta gtgaaattga attagtattc aggcctcatc ccacacttat ggaaaaagat
 721 gacagtgcac agacgagata cataaagact tctggtaacg ccactgttga tcacttatcc
 781 aagtatctgg ctgtgaggtt agctttagaa gaacttcgaa gcaaaggtga atcaaaccag
 841 atgaaccttg atacagccag tgagaagcag tataccattt atatagcaac agccagtggc
 901 cagttcactg tattaaatgg ctctttttct ttggaattgg tcagtgagaa atactggaaa
 961 gtgaacaaac ccatggaact ttattacgca cctacaaagg agcacaaatg a
```

//

30. RNF2 (*Homo sapiens*) Protein sequence

```
   1 msqavqtngt qplsktwels lyelqrtpqe aitdgleivv sprslhselm cpicldmlkn
  61 tmttkeclhr fcadciital rsgnkecptc rkklvskrsl rpdpnfdali skiypsrdey
 121 eahqervlar inkhnnqqal shsieeglki qamnrlqrgk kqqiengsga edngdsshcs
 181 nasthsnqea gpsnkrtkts ddsgleldnn naamaidpvm dgaseielvf rphptlmekd
 241 dsaqtryikt sgnatvdhls kylavrlale elrskgesnq mnldtasekq ytiyiatasg
 301 qftvlngsfs lelvsekywk vnkpmelyya ptkehk
//
```

31. RNF2 (*Mus musculus*) cDNA sequence

```
   1 atgtctcagg ctgtgcagac aaatggaact caaccattaa gcaaaacatg ggaactcagt
  61 ttgtatgagt tacaacgaac acctcaggag gcaataacag atggcttgga aattgtggtt
 121 tcacctagaa gtctacacag tgaattaatg tgcccaattt gtttggatat gttaaagaac
 181 accatgacta caaaggagtg tttacatcgg ttttgcgcgg attgtattat cacagccctt
 241 agaagtggca acaaagagtg tcctacctgt cggaaaaaac tggtttctaa aagatcacta
 301 aggccagacc cgaactttga tgcactcatc agcaagattt atcccagtcg tgatgagtat
 361 gaagcgcatc aggaaagggt cttagcaagg atcaacaaac acaacaatca gcaggctctc
 421 agccacagca tcgaggaggg gctgaagata caggccatga acagattaca gcgaggcaaa
 481 aagcagcaga tagaaaatgg tagtggagca gaagataatg gtgacagctc ccactgtagt
 541 aacgcatcca cacacagcaa ccaggaagcg ggcccgagta acaaacggac caaaacctct
 601 gatgactctg ggcttgaact tgataacaac aatgcagcag tggcgattga tccagtcatg
 661 gacggtgcca gtgagattga gttagtcttc aggccccatc caactcttat ggaaaaggac
 721 gacagcgcac agacaagata cataaagact tcaggcaatg ccactgttga tcacttatcc
 781 aagtatctgg ctgtgaggtt agctttagaa gaacttcgaa gcaaaggaga atcaaaccag
 841 atgaacctgg atacagccag tgagaagcag tacaccattt acatagccac agccagtggc
 901 cagttcaccg ttttaaatgg ctccttttct ttggaattgg tcagtgagaa atactggaaa
 961 gtgaacaaac ccatggaact ttattatgca cccaccaagg agcacaaatg a
//
```

32. RNF2 (*Mus musculus*) Protein sequence

```
   1 msqavqtngt qplsktwels lyelqrtpqe aitdgleivv sprslhselm cpicldmlkn
  61 tmttkeclhr fcadciital rsgnkecptc rkklvskrsl rpdpnfdali skiypsrdey
 121 eahqervlar inkhnnqqal shsieeglki qamnrlqrgk kqqiengsga edngdsshcs
 181 nasthsnqea gpsnkrtkts ddsgleldnn naavaidpvm dgaseielvf rphptlmekd
 241 dsaqtryikt sgnatvdhls kylavrlale elrskgesnq mnldtasekq ytiyiatasg
 301 qftvlngsfs lelvsekywk vnkpmelyya ptkehk
//
```

33. ANK3 (*Homo sapiens*) cDNA sequence

```
   1 atggctcatg cagcctcaca attaaagaaa aacagggatt tagaaatcaa tgctgaagaa
  61 gagcctgaga aaaaaaggaa acaccgcaaa cggtcccggg atcggaagaa aaagtctgat
 121 gccaatgcaa gttacttaag agcagctcga gctggacacc ttgaaaaggc cctcgactac
 181 ataaaaaaatg gagttgacat caacatttgc aatcagaatg ggttgaacgc tctccacctt
 241 gcttccaaag aaggccatgt agaggttgtt tctgagctgc tgcagagaga agccaatgtg
 301 gatgcagcta caaagaaagg aaacacagca ttgcacatcg catctttggc tgggcaagca
```

361 gaggtggtaa aagtcttggt tacaaatgga gccaatgtca atgcacaatc tcagaatggt
421 ttcacgccat tgtatatggc agcccaggaa aatcacctgg aagttgtcaa gtttcttctt
481 gacaatggtg caagccagag cctagccaca gaggatggct tcacaccatt ggcagtggct
541 ttgcaacaag gtcacgacca agtcgtttcg ctcctgctag agaatgacac caaaggaaaa
601 gtgcgtctcc cagctcttca tatcgcggcc cgaaaagacg acacgaaagc cgccgccctg
661 ctgctgcaga atgacaacaa tgcagatgtg gaatcaaaga gtggcttcac tccgctccac
721 atagctgctc actatggaaa tatcaatgta gccacgttgc tgttaaaccg agcggctgct
781 gtggatttca ccgcaaggaa tgacatcact cctttacatg ttgcatcaaa aagaggaaat
841 gcaaatatgg taaaactatt gctcgatcga ggagctaaaa tcgatgccaa aaccagggat
901 ggtctgacac cactgcactg tggagcaagg agtggccacg agcaggtggt agaaatgttg
961 cttgatcgag ctgcccccat tctttcaaaa accaagaatg gattatctcc attgcacatg
1021 gccacacaag gggatcattt aaactgcgtc cagcttctcc tccagcataa tgtacccgtg
1081 gatgatgtca ccaatgacta cctgactgcc ctacacgtgg ctgcccactg tggccattac
1141 aaagttgcca aggttctctt ggataagaaa gctaacccca atgccaaagc cctgaatggc
1201 tttacccctc ttcatattgc ctgcaagaag aatcgaatta aagtaatgga actccttctg
1261 aaacacggtg catccatcca agctgtaacc gagtcgggcc ttaccccaat ccatgttgct
1321 gccttcatgg ggcatgtaaa tattgtatca caactaatgc atcatggagc ctcaccaaac
1381 accaccaatg tgagaggaga aacagcactg cacatggcag ctcgctccgg ccaagctgaa
1441 gttgtgcggt atctggtaca agacggagct caggtagaag ctaaagctaa ggatgaccaa
1501 acaccactcc acatttcagc ccgactgggg aaagcagaca tagtacaaca gctgttgcag
1561 caaggggcat ctccaaatgc agccacaact tctgggtaca ccccacttca cctttccgcc
1621 cgagaggggc atgaggatgt ggccgcgttc cttttggatc atggagcgtc tttatctata
1681 acaacaaaga aaggatttac tcctcttcat gtggcagcaa aatatggaaa gcttgaagtc
1741 gccaatctcc tgctacagaa aagtgcatct ccagatgctg ctgggaagag cgggctaaca
1801 ccactgcatg tagctgcaca ttacgataat cagaaagtgg cccttctgct tttggaccaa
1861 ggagcctcac ctcacgcagc cgcaaagaat ggttatacgc cactgcacat cgctgccaaa
1921 aagaaccaga tggacatagc gacaactctg ctggaatatg gtgctgatgc caacgcagtt
1981 acccggcaag gaattgcttc cgtccatctc gcagctcagg aagggcacgt ggacatggtg
2041 tcgctgctcc tcggtagaaa tgcgaatgtg aacctgagca ataagagcgg cctgacccca
2101 ctccatttgg ctgctcaaga agatcgagtg aatgtggcag aagtcctcgt aaaccaaggg
2161 gctcatgtgg acgcccagac aaagatggga tacacaccac tgcatgtggg ctgccactat
2221 ggaaatatca agattgttaa tttcctgctc cagcattctg caaaagttaa tgccaaaaca
2281 aagaatgggt atacgccatt acatcaagca gcacagcagg ggcatacgca tataataaat
2341 gtcttacttc agaacaacgc ctcccccaat gaactcactg tgaatgggaa tactgccctt
2401 ggcattgccc ggcgcctcgg ctacatctca gtagtggaca ccctgaagat agtgaccgaa
2461 gagaccatga ccacaactac tgtcacagag aagcacaaaa tgaatgttcc agaaacgatg
2521 aatgaagttc ttgatatgtc tgatgatgaa gttcgtaaag ccaatgcccc tgaaatgctc
2581 agtgatggcg aatatatctc agatgttgaa gaaggtgaag atgcaatgac cggggacaca
2641 gacaaatatc ttgggccaca ggaccttaag gaattgggtg atgattccct gcctgcagag
2701 ggttacatgg gctttagtct cggagcgcgt tctgccagcc tccgctcctt cagttcggat
2761 aggtcttaca ccttgaacag aagctcctat gcacgggaca gcatgatgat tgaagaactc
2821 cttgtgccat ccaaagagca gcatctaaca ttcacaaggg aatttgattc agattctctt
2881 agacattaca gctgggctgc agacacctta gacaatgtca atcttgtttc aagccccatt
2941 cattctgggt ttctggttag ctttatggtg gacgcgagag ggggctccat gagaggaagc

3001 cgtcatcacg ggatgagaat catcattcct ccacgcaagt gtacggcccc cactcgaatc
3061 acctgccgtt tggtaaagag acataaactg gccaacccac cccccatggt ggaaggagag
3121 ggattagcca gtaggctggt agaaatgggt cctgcagggg cacaattttt aggccctgtc
3181 atagtggaaa tccctcactt tgggtccatg agaggaaaag agagagaact cattgttctt
3241 cgaagtgaaa atggtgaaac ttggaaggag catcagtttg acagcaaaaa tgaagattta
3301 accgagttac ttaatggcat ggatgaagaa cttgatagcc cagaagagtt agggaaaaag
3361 cgtatctgca ggattatcac gaaagatttc ccccagtatt ttgcagtggt ttcccggatt
3421 aagcaggaaa gcaaccagat tggtcctgaa ggtggaattc tgagcagcac cacagtgccc
3481 cttgttcaag catctttccc agagggtgcc ctaactaaaa gaattcgagt gggcctccag
3541 gcccagcctg ttccagatga aattgtgaaa aagatccttg gaaacaaagc aacttttagc
3601 ccaattgtca ctgtggaacc aagaagacgg aaattccata aaccaatcac aatgaccatt
3661 ccggtgcccc cgccctcagg agaaggtgta tccaatggat acaaagggga cactacaccc
3721 aatctgcgtc ttctctgtag cattacaggg ggcacttcgc ctgctcagtg ggaagacatc
3781 acaggaacaa ctcctttgac gtttataaaa gattgtgtct cctttacaac caatgtttca
3841 gccagatttt ggcttgcaga ctgccatcaa gttttagaaa ctgtggggtt agccacgcaa
3901 ctgtacagag aattgatatg tgttccatat atggccaagt ttgttgtttt tgccaaaatg
3961 aatgatcccg tagaatcttc cttgcgatgt ttctgcatga cagatgacaa agtggacaaa
4021 actttagagc aacaagagaa ttttgaggaa gtcgcaagaa gcaaagatat tgaggttctg
4081 gaaggaaaac ctatttatgt tgattgttat ggaaatttgg ccccacttac caaaggagga
4141 cagcaacttg tttttaactt ttattctttc aaagaaaata gactgccatt ttccatcaag
4201 attagagaca ccagccaaga gccctgtggt cgtctgtctt ttctgaaaga accaaagaca
4261 acaaaaggac tgcctcaaac agcggtttgc aacttaaata tcactctgcc agcacataaa
4321 aaggagacag agtcagatca agatgatgag attgagaaaa cagatagacg acagagcttc
4381 gcatccttag ctttacgtaa gcgctacagc tacttgactg agcctggaat gattgaacgg
4441 agtacaggag caacaagatc cctccccacc acttactcat acaagccatt cttttctaca
4501 agaccatacc agtcctggac aacagctccg attacagtgc ctgggccagc caagtcaggc
4561 ttcacttcct tatcaagttc ttcctctaat acgccatcag cttctccgtt aaaatcaata
4621 tggtctgttt cgacaccttc tccaatcaaa tccacattag gcgcgtcaac tacatcttca
4681 gttaaatcca ttagtgacgt ggcatctcca attagatcct ttcggacaat gtcttcgccg
4741 ataaaaactg tggtgtcaca atctccatac aatatccaag tttcctctgg taccctggct
4801 agagctccag cagtcacgga agctacgccc ttaaaagggc tggcatccaa ttctacgttt
4861 tcctctcgaa cctctccagt gactacagca gggtctcttt tggagaggtc atcaattact
4921 atgacacccc ctgcctcccc caaatcaaac attaatatgt attcctcaag tttgccattt
4981 aagtcaatta ttacatcagc agcaccgcta atatcttcac ctttaaagtc agtggtgtct
5041 ccagttaaat cagcagttga tgtcatttca tcagccaaaa ttacaatggc atcttctctc
5101 tcatcacctg tgaagcagat gcctggacat gcagaggtag cattagtcaa tggatctatt
5161 tcccctctaa aatatccatc atcctcaact ttaattaatg gatgcaaagc cactgccacg
5221 ttacaggaaa aaatttcttc tgctacaaac tctgtgagct ctgtggtcag tgcagccact
5281 gacacagttg agaaagtgtt ttctaccacg actgcaatgc catttcccc actcaggtca
5341 tatgtttctg cagcaccatc agcttttcag tctctaagaa ctccttccgc aagtgcactc
5401 tatacatccc ttgggtcgtc aatatctgca actacctcat ctgtaacttc atcaattata
5461 acagtgccag tatactctgt agtcaatgtt ttgccagaac cagcattaaa gaaacttcca
5521 gactctaatt catttacaaa atcagcagca gccttgctgt cacccattaa aacattgact
5581 acggagacac atcctcagcc tcacttcagt cgaacttcat ctccagttaa gtcatctttg

5641 ttccttgcac cctctgccct taagttgtct acaccatctt ctttatcttc cagtcaggag
5701 atactaaaag atgtagctga aatgaaagag gacctaatgc ggatgaccgc aatactacag
5761 acagatgtgc ctgaggagaa gccattccaa cctgaactcc caaaggaagg gagaatagat
5821 gatgaagaac ctttcaaaat tgtagagaaa gtaaaggaag acttagtgaa agttagtgaa
5881 atccttaaaa aggatgtatg tgtagataat aaaggatcac ccaaatcacc aaagagtgac
5941 aaaggacact ctcctgaaga tgactggata gaatttagtt cggaagaaat ccgggaagcc
6001 agacaacaag ctgctgcgag ccagtctcca tctctgccag agagagtgca agtaaaagca
6061 aaagccgcct ccgaaaagga ttataacttg accaaagtta ttgattacct aacaaatgat
6121 attgggagta gttcactgac aaacttaaaa tacaagtttg aggatgcaaa gaaggatggt
6181 gaggagagac agaaaagagt tttaaaacca gcaattgctt tgcaggaaca caaactcaaa
6241 atgcctccag cctccatgag gacttccacc tctgagaaag aattgtgtaa aatggctgat
6301 tccttttttg gaacagatac tattttagag tctcctgatg acttttctca acacgaccaa
6361 gataaaagtc ccttgtctga cagtggcttt gaaacaagaa gtgaaaagac accttcagcc
6421 ccacaaagcg ctgaaagcac tggtcctaaa ccacttttc atgaagttcc catccctcct
6481 gtcattacag aaacaagaac tgaagtggtt catgttatca ggagctatga tccctcagct
6541 ggggatgttc cccagaccca accagaggag cctgtgtcac ctaaaccttc acctactttt
6601 atggaattgg aaccaaagcc caccacctct agtattaaag aaaaggttaa agcatttcaa
6661 atgaaagcca gtagtgaaga agatgaccac aatcgggttt taagcaaagg catgcgtgtt
6721 aaagaagaga ctcacataac cacaaccacc agaatggttt atcattctcc accaggcggt
6781 gaaggtgcat ctgaaagaat tgaagaaacc atgtcagtcc atgacatcat gaaggccttt
6841 cagtccgggc gggatccttc caaagaactg gcaggtctgt ttgaacataa gtcggcagtg
6901 tctccagatg ttcacaagtc tgctgctgaa acctcagccc agcatgcaga gaaggacaac
6961 caaatgaaac ccaaactgga gcgtataata gaagtccaca tcgaaaaagg taaccaagct
7021 gagcccactg aagtcattat tagagaaacc aaaaagcatc cagaaaaaga aatgtatgta
7081 tatcagaaag acttatcccg gggagatatt aacctaaaag attttctgcc agaaaaacac
7141 gatgcttttc cttgttcaga ggaacagggt cagcaagaag aagaagaact tactgctgaa
7201 gagtcattgc cttcttatct ggagtcttcc agagtaaaca ctcctgtgtc ccaagaagaa
7261 gatagccgcc ctagttctgc tcaactcata tctgatgact cttataaaac attgaagctt
7321 ttgagtcaac actcaataga ataccatgac gatgagttgt cagaactaag aggggagtct
7381 tacaggtttg ctgagaaaat gcttctgtca gaaaagctag atgtgtctca ttctgatact
7441 gaggaatcgg ttacagacca tgcaggaccc cctagctcag agttacaggg gtctgataag
7501 cggtccagag aaaaaatagc cactgccccc aaaaaagaaa ttctctccaa aatctataaa
7561 gatgtttctg aaaatggtgt aggtaaagtg tctaaagatg agcattttga taaagtgaca
7621 gtgttgcact attctggcaa tgttagtagt ccaaaacatg ccatgtggat gcgctttact
7681 gaggacagat tagacagagg tagagagaag ttgatatatg aagatagggt ggacaggact
7741 gtgaaggagg ctgaagaaaa actgactgaa gtgtcacagt tttttcgtga caaaactgaa
7801 aagctaaatg atgaactgca gtccccagag aaaaaggcac gccctaaaaa tggcaaagaa
7861 tattcttctc aaagccctac cagtagcagc cctgagaaag tgctactgac agaactgctg
7921 gcatccaatg atgagtgggt taaggcaaga cagcatggcc ctgatggaca aggcttcccc
7981 aaggccgagg agaaggcacc cagtctgccc agcagcccag agaagatggt tctctcccaa
8041 cagactgagg acagcaagtc cacagtggaa gccaaaggaa gtatttcaca gagcaaagca
8101 ccagatgggc cccagtctgg attccagctc aaacaatcta aactcagttc cattagatta
8161 aaatttgaac aaggcacaca cgcaaaaagt aaggacatgt ctcaagaaga cagaaagtca
8221 gatggccagt ccagaatccc agttaaaaaa atacaggaga gcaagctacc cgtctaccaa

8281 gtttttgcta gagaaaaaca gcagaaggcc atagacctcc cagatgaaag tgtatctgtg
8341 caaaaagatt ttatggtatt aaaaaccaaa gatgagcatg cccaaagcaa cgaaattgtt
8401 gtaaatgatt ctggctctga taatgtgaaa aaacagagaa ctgaaatgtc aagtaaagca
8461 atgcctgact ctttttctga gcagcaggct aaagacttgg catgtcatat aacctcagat
8521 ttagcaacta ggggaccatg ggacaaaaag gtctttagaa catgggagag ttcgggagcc
8581 actaacaata agtctcagaa agaaaaactt tcgcatgtac ttgttcatga tgtaagagag
8641 aatcacattg gtcaccctga gagtaaaagt gttgatcaaa agaatgaatt tatgtctgtg
8701 actgagagag aacgcaaatt gttaacaaac ggctctctct cagaaattaa agaaatgact
8761 gtaaaatctc cctccaaaaa agtcttatat agggaatatg ttgtgaaaga aggggaccat
8821 ccaggcggat tgcttgatca gccttccagg aggagcgaga gctcagcagt gtcacacatt
8881 cccgtcagag ttgctgatga gaggagaatg ctgtcttcta atattcccga tggtttttgt
8941 gaacagtcgg catttccaaa acatgaacta tcacaaaaat tgtcccagtc aagcatgagt
9001 aaagagacag ttgagacaca gcactttaat tctatagaag atgaaaaagt tacctattca
9061 gaaatcagca aagtttccaa acaccagagt tatgtaggtt tatgcccacc tctcgaggaa
9121 accgaaacct cccccaccaa atctcctgat tctttagagt ttagcccagg aaaggaatct
9181 ccctctagtg atgtattcga ccacagtccc attgatggat tggaaaaact cgcaccacta
9241 gcccagacag agggagggaa agagataaaa actttacccg tttatgtcag ttttgtacaa
9301 gtggggaagc aatatgaaaa ggagatacaa caaggaggtg taaaaaaaat cataagtcag
9361 gaatgtaaga cagtacaaga aaccaggggg acctttttata caactagaca gcaaaagcaa
9421 cctccttctc cccaaggtag tccagaagat gatactctag agcaagtatc ctttctagac
9481 agctctggga aaagcccttt aaccccagaa acacccagtt cagaggaagt gagttatgaa
9541 tttacatcta agacacctga ctcgctcata gcttatatac caggcaaacc cagcccaatt
9601 cccgaggttt ctgaggagtc agaggaggag gaacaggcca agtcaacctc ccttaagcag
9661 actacagtgg aggaaacagc agttgagcgt gaaatgccta atgacgtgag caaagactct
9721 aaccaaagac ccaaaaataa cagagttgcc tatattgaat ttccccctcc tccaccactg
9781 gatgcggacc agattgagtc agataagaag catcattatc tcccagaaaa agaggttgac
9841 atgattgaag tcaatctgca agatgagcat gacaagtacc agctggctga acctgtcatt
9901 agagtgcagc caccttcacc agttcctccc ggggcagacg tcagtgattc aagcgatgac
9961 gaatctattt atcagccagt cccagttaaa aaatatacct tcaaattaaa ggaagtggac
10021 gatgaacaaa aagaaaaacc caaagcttct gctgaaaagg cttccaacca gaaagaactg
10081 gaaagtaatg gatctggaaa agataatgaa tttggccttg gccttgattc acctcagaat
10141 gaaattgccc agaatgggaa caacgaccag tccatcacag agtgttccat tgccaccaca
10201 gcagagtttt ctcatgacac ggatgccaca gagatcgact ctctggatgg ctatgacctg
10261 caagatgaag atgatggctt gacagagagt gattctaaac tcccaattca agccatggaa
10321 attaagaaag atatctggaa cacagagggc attctgaagc cagctgaccg ctcttttagc
10381 caaagtaaac ttgaagttat cgaggaggag ggaaaggtgg gaccagatga ggacaagcca
10441 ccttctaaaa gttcttcatc tgaaaagact cctgataaga ctgatcagaa gtcaggggcc
10501 cagttcttca cactggaagg cagacatcct gacagatcag tgtttcctga tacttacttc
10561 agttacaaag tagatgaaga atttgccact ccttttaaaa cagtagctac caaaggtcta
10621 gattttgacc cttggtctaa taaccgaggg gatgatgaag tttttgacag taaatcacgg
10681 gaagatgaaa ctaagccatt tgggctggcg gtagaagacc gctctccagc aacaacccct
10741 gatacaacgc cagccagaac gccaactgat gaaagtaccc caactagtga gcctaacccc
10801 ttcccatttc atgaaggaaa aatgtttgag atgactcgca gtggtgcaat tgacatgagc
10861 aagagggatt ttgttgaaga gaggctccaa tttttccaga ttggtgagca tacttctgaa

10921 gggaagtcag gggaccaggg ggaaggggat aaaagtatgg tcactgccac accacagcca
10981 cagtcagggg acaccactgt agaaaccaat ctagagagaa atgtagagac acctacagtg
11041 gaacctaacc ccagcatccc gaccagcgga gagtgtcagg aaggcacatc cagtagtggc
11101 tccctggaga aatcagcagc agccactaac acctctaaag ttgaccccaa gttgcgcacg
11161 cctataaaaa tgggaatttc tgcatccacc atgaccatga agaaagaagg ccctggagaa
11221 ataacagata agatagaagc ggtgatgacc agttgtcagg gattagaaaa tgaaactata
11281 acaatgattt caaatacagc caatagccag atgggcgtta ggccccatga aaaacatgat
11341 tttcaaaaag ataactttaa taacaacaac aatttggatt cttccactat acagacagat
11401 aacattatga gtaatatagt tctgacagaa cattctgcac ccacttgtac cacagagaaa
11461 gataacccag tgaaagtctc atcaggaaaa aagacagggg tactacaagg acactgtgta
11521 agagataagc agaaagttct tggagaacag caaaaaacaa aggaattgat agggattagg
11581 caaaaatcca aacttcccat aaaggccact tcaccaaaag ataccttccc accgaaccat
11641 atgtcaaaca ctaaagcaag taaaatgaag caggttagtc aatccgagaa aaccaaagcc
11701 cttactactt cttcatgtgt agatgtaaag tccagaattc cagtgaaaaa cacacacagg
11761 gataacataa ttgcagttag aaaagcatgt gccacacaaa agcaagggca gccagagaaa
11821 ggcaaggcca aacagcttcc atccaagttg ccagtaaagg taagatccac ctgtgtcact
11881 accaccacca ccactgccac caccaccacc actaccacca ctaccaccac caccagctgc
11941 acagttaaag ttaggaaaag tcagctcaag gaagtatgta aacattccat tgaatatttt
12001 aagggaatta gtggtgagac cttaaagctt gtggaccgcc tctctgaaga agaaaaaaag
12061 atgcagtccg agttgtccga tgaggaagaa agtacctcaa gaaacacgtc gttgtccgag
12121 acttcccggg gtggccagcc ttcggttaca acgaagtctg ctagagataa gaaaacagag
12181 gcagcacctt taaaatcaaa gagtgaaaag gccggcagtg agaaaaggag cagtagaagg
12241 actggtccac agagtccatg tgaacggaca gatatcagga tggcaatagt agccgatcac
12301 ctgggactta gttggacaga actggcaagg gaactgaatt tttcagtgga tgaaatcaat
12361 caaatacgtg tggaaaatcc aaattcttta atttctcaga gcttcatgtt attaaaaaaa
12421 tgggttacca gagacggaaa aaatgccaca actgatgcct taacttcggt cttgacaaaa
12481 attaatcgaa tagatatagt gacactgcta gaaggaccaa tatttgatta tggaaatatt
12541 tcaggcacca gaagttttgc agatgagaac aatgttttcc atgaccctgt tgatggttgg
12601 cagaatgaga catcaagtgg aaacctagag tcctgcgctc aagctcgaag agtaactggt
12661 gggttactag atcgactgga tgacagccct gaccagtgta gagattccat tacctcatat
12721 ctcaaaggag aagctggcaa atttgaagca aatggaagcc atacagaaat cactccagaa
12781 gcaaagacaa aatcttactt tccagaatcc caaaatgatg taggaaaaca gagtaccaag
12841 gaaactctga aaccaaaaat acatggatct ggtcatgttg aagaaccagc atcaccacta
12901 gcagcatatc agaaatctct agaagaaacc agcaagctta taatagaaga gactaaaccc
12961 tgtgtgcctg tcagtatgaa aaagatgagt aggacttctc cagcagatgg caagccaagg
13021 cttagcctcc atgaagaaga ggggtccagt gggtctgagc aaaagcaggg agaaggtttt
13081 aaggtgaaaa cgaagaaaga aatccggcat gtggaaaaga agagccactc gtaa

//

34. ANK3 (*Homo sapiens*) Protein sequence

1 mahaasqlkk nrdleinaee epekkrkhrk rsrdrkkksd anasylraar aghlekaldy
61 ikngvdinic nqnglnalhl askeghvevv sellqreanv daatkkgnta lhiaslagqa
121 evvkvlvtng anvnaqsqng ftplymaaqe nhlevvkfll dngasqslat edgftplava
181 lqqghdqvvs lllendtkgk vrlpalhiaa rkddtkaaal llqndnnadv esksgftplh
241 iaahygninv atlllnraaa vdftarndit plhvaskrgn anmvkllldr gakidaktrd

301 gltplhcgar sgheqvveml ldraapilsk tknglsplhm atqgdhlncv qlllqhnvpv
361 ddvtndylta lhvaahcghy kvakvlldkk anpnakalng ftplhiackk nrikvmelll
421 khgasiqavt esgltpihva afmghvnivs qlmhhgaspn ttnvrgetal hmaarsgqae
481 vvrylvqdga qveakakddq tplhisarlg kadivqqllq qgaspnaatt sgytplhlsa
541 reghedvaaf lldhgaslsi ttkkgftplh vaakygklev anlllqksas pdaagksglt
601 plhvaahydn qkvalllldq gasphaaakn gytplhiaak knqmdiattl leygadanav
661 trqgiasvhl aaqeghvdmv slllgrnanv nlsnksgltp lhlaaqedrv nvaevlvnqg
721 ahvdaqtkmg ytplhvgchy gnikivnfll qhsakvnakt kngytplhqa aqqghthiin
781 vllqnnaspn eltvngntal giarrlgyis vvdtlkivte etmttttvte khkmnvpetm
841 nevldmsdde vrkanapeml sdgeyisdve egedamtgdt dkylgpqdlk elgddslpae
901 gymgfslgar saslrsfssd rsytlnrssy ardsmmieel lvpskeqhlt ftrefdsdsl
961 rhyswaadtl dnvnlvsspi hsgflvsfmv darggsmrgs rhhgmriiip prkctaptri
1021 tcrlvkrhkl anpppmvege glasrlvemg pagaqflgpv iveiphfgsm rgkerelivl
1081 rsengetwke hqfdsknedl tellngmdee ldspeelgkk ricriitkdf pqyfavvsri
1141 kqesnqigpe ggilssttvp lvqasfpega ltkrirvglq aqpvpdeivk kilgnkatfs
1201 pivtveprrr kfhkpitmti pvpppsgegv sngykgdttp nlrllcsitg gtspaqwedi
1261 tgttpltfik dcvsfttnvs arfwladchq vletvglatq lyrelicvpy makfvvfakm
1321 ndpvesslrc fcmtddkvdk tleqqenfee varskdievl egkpiyvdcy gnlapltkgg
1381 qqlvfnfysf kenrlpfsik irdtsqepcg rlsflkepkt tkglpqtavc nlnitlpahk
1441 ketesdqdde iektdrrqsf aslalrkrys yltepgmier stgatrslpt tysykpffst
1501 rpyqswttap itvpgpaksg ftslsssssn tpsasplksi wsvstpspik stlgasttss
1561 vksisdvasp irsfrtmssp iktvvsqspy niqvssgtla rapavteatp lkglasnstf
1621 ssrtspvtta gsllerssit mtppaspksn inmyssslpf ksiitsaapl issplksvvs
1681 pvksavdvis sakitmassl sspvkqmpgh aevalvngsi splkypssst lingckatat
1741 lqekissatn svssvvsaat dtvekvfstt tampfsplrs yvsaapsafq slrtpsasal
1801 ytslgssisa ttssvtssii tvpvysvvnv lpepalkklp dsnsftksaa allspiktlt
1861 tethpqphfs rtsspvkssl flapsalkls tpsslsssqe ilkdvaemke dlmrmtailq
1921 tdvpeekpfq pelpkegrid deepfkivek vkedlvkvse ilkkdvcvdn kgspkspksd
1981 kghspeddwi efsseeirea rqqaaasqsp slpervqvka kaasekdynl tkvidyltnd
2041 igsssltnlk ykfedakkdg eerqkrvlkp aialqehklk mppasmrtst sekelckmad
2101 sffgtdtile spddfsqhdq dksplsdsgf etrsektpsa pqsaestgpk plfhevpipp
2161 vitetrtevv hvirsydpsa gdvpqtqpee pvspkpsptf melepkptts sikekvkafq
2221 mkasseeddh nrvlskgmrv keethitttt rmvyhsppgg egaserieet msvhdimkaf
2281 qsgrdpskel aglfehksav spdvhksaae tsaqhaekdn qmkpklerii evhiekgnqa
2341 epteviiret kkhpekemyv yqkdlsrgdi nlkdflpekh dafpcseeqg qqeeeeltae
2401 eslpsyless rvntpvsqee dsrpssaqli sddsyktlkl lsqhsieyhd delselrges
2461 yrfaekmlls ekldvshsdt eesvtdhagp psselqgsdk rsrekiatap kkeilskiyk
2521 dvsengvgkv skdehfdkvt vlhysgnvss pkhamwmrft edrldrgrek liyedrvdrt
2581 vkeaeeklte vsqffrdkte klndelqspe kkarpkngke yssqsptsss pekvlltell
2641 asndewvkar qhgpdgqgfp kaeekapslp sspekmvlsq qtedskstve akgsisqska
2701 pdgpqsgfql kqsklssirl kfeqgthaks kdmsqedrks dgqsripvkk iqesklpvyq
2761 vfarekqqka idlpdesvsv qkdfmvlktk dehaqsneiv vndsgsdnvk kqrtemsska
2821 mpdsfseqqa kdlachitsd latrgpwdkk vfrtwessga tnnksqkekl shvlvhdvre
2881 nhighpesks vdqknefmsv tererkllln gslseikemt vkspskkvly reyvvkegdh

2941 pgglldqpsr rsessavshi pvrvaderrm lssnipdgfc eqsafpkhel sqklsqssms
3001 ketvetqhfn siedekvtys eiskvskhqs yvglcpplee tetsptkspd slefspgkes
3061 pssdvfdhsp idgleklapl aqteggkeik tlpvyvsfvq vgkqyekeiq qggvkkiisq
3121 ecktvqetrg tfyttrqqkq ppspqgsped dtleqvsfld ssgkspltpe tpsseevsye
3181 ftsktpdsli ayipgkpspi pevseeseee eqakstslkq ttveetaver empndvskds
3241 nqrpknnrva yiefpppppl dadqiesdkk hhylpekevd mievnlqdeh dkyqlaepvi
3301 rvqppspvpp gadvsdssdd esiyqpvpvk kytfklkevd deqkekpkas aekasnqkel
3361 esngsgkdne fglgldspqn eiaqngnndq sitecsiatt aefshdtdat eidsldgydl
3421 qdeddgltes dsklpiqame ikkdiwnteg ilkpadrsfs qsklevieee gkvgpdedkp
3481 psksssssekt pdktdqksga qfftlegrhp drsvfpdtyf sykvdeefat pfktvatkgl
3541 dfdpwsnnrg ddevfdsksr edetkpfgla vedrspattp dttpartptd estptsepnp
3601 fpfhegkmfe mtrsgaidms krdfveerlq ffqigehtse gksgdqgegd ksmvtatpqp
3661 qsgdttvetn lernvetptv epnpsiptsg ecqegtsssg sleksaaatn tskvdpklrt
3721 pikmgisast mtmkkegpge itdkieavmt scqgleneti tmisntansq mgvrphekhd
3781 fqkdnfnnnn nldsstiqtd nimsnivlte hsaptcttek dnpvkvssgk ktgvlqghcv
3841 rdkqkvlgeq qktkeligir qksklpikat spkdtfppnh msntkaskmk qvsqsektka
3901 lttsscvdvk sripvknthr dniiavrkac atqkqgqpek gkakqlpskl pvkvrstcvt
3961 tttttatttt tttttttttsc tvkvrksqlk evckhsieyf kgisgetlkl vdrlseeekk
4021 mqselsdeee stsrntslse tsrggqpsvt tksardkkte aaplksksek agsekrssrr
4081 tgpqspcert dirmaivadh lglswtelar elnfsvdein qirvenpnsl isqsfmllkk
4141 wvtrdgknat tdaltsvltk inridivtll egpifdygni sgtrsfaden nvfhdpvdgw
4201 qnetssgnle scaqarrvtg glldrlddsp dqcrdsitsy lkgeagkfea ngshteitpe
4261 aktksyfpes qndvgkqstk etlkpkihgs ghveepaspl aayqksleet skliieetkp
4321 cvpvsmkkms rtspadgkpr lslheeegss gseqkqgegf kvktkkeirh vekkshs

//

35. ANK3 (*Mus musculus*) cDNA sequence

1 atgagtgaag agccaaagga gaagcccgcc aagcctgctc ataggaagag gaaaggaaaa
61 aagtctgatg ccaacgcaag ttacttaaga gcagctcggg cagggcacct ggaaaaggcc
121 cttgactaca tcaaaaatgg agtggacgtc aacatctgta accagaatgg attgaatgca
181 ctccatcttg cttccaaaga aggccatgtg gaagtggtct ctgagctgct gcagagggaa
241 gccaatgttg atgccgccac aaagaaagga aacacggcct tacacatcgc atctttggct
301 gggcaagcgg aagtggtcaa ggtcttggtt acgaacggag cgaatgtcaa cgcacaatct
361 cagaatggct tcacaccatt gtatatggca gcccaggaga accacctgga agtcgtcagg
421 tttcttctgg acaatggcgc cagccaaagc ctggccacag aggacggctt cacgccattg
481 gccgtggctc tgcaacaagg tcatgaccaa gtcgtgtccc tcctgctcga gaacgacacg
541 aagggaaaag tgcgcctccc agccctccac atcgcagccc ggaaagacga caccaaggca
601 gcagctctgc tcctgcagaa tgacacaaac gcggacgtgg agtcaaagag tggcttcacc
661 ccgctccaca tagctgccca ctatgggaac atcaatgtgg ccacgttgct gttaaaccga
721 gcggctgctg tggacttcac cgcacggaat gacatcactc ccttacacgt tgcctcgaag
781 cgaggaaatg caaatatggt gaagctattg ctggaccggg gtgcgaagat cgatgccaag
841 accagggacg gtctgactcc gttgcactgt ggggcgagaa gtggccatga gcaggtggta
901 gagatgttgc ttgacagatc cgcccccatc ctttcaaaaa ccaagaatgg attgtcgcca
961 ctgcacatgg ccacacaagg agaccattta aactgcgtcc aactcctcct ccagcacaac
1021 gtgcccgtgg acgacgtcac caacgactac ctgactgccc tccatgtggc tgcccactgc

1081 ggccattaca aagttgccaa ggttcttttg gataagaaag ctagccccaa tgccaaagcc
1141 ctgaatggct tcacccctct ccatatcgcc tgcaaaaaga accgcatccg agtaatggaa
1201 ctccttttga agcacggtgc atctattcaa gccgtaaccg agtcgggcct taccccaatc
1261 catgttgctg ccttcatggg acatgtaaat atcgtgtcac agctaatgca tcatggagcc
1321 tccccaaaca ccaccaatgt gagaggagag acggcattgc atatggcggc tcggtccgga
1381 caagcagaag tggtgcggta tctggtccaa gatggggctc aggtagaagc aaaagctaag
1441 gatgaccaga ctccactcca catctcagcc cgacttggga aagctgacat agtgcaacaa
1501 ctgttacagc aaggagcatc ccccaatgca gcaacaactt ctgggtacac ccccttcac
1561 cttgcggcca gagaggggca tgaggatgta gctgcgttcc tcctggatca tggagcatct
1621 ttatccataa caacaaagaa gggattcacc cctctgcacg tggcagccaa atacggaaag
1681 cttgaagtcg caagtctcct gctgcagaag agtgcgtctc ccgatgccgc agggaagagc
1741 gggctaactc cactgcatgt agcagcgcat tacgataatc agaaagtggc ccttctgctc
1801 ttggaccagg gagcctcacc ccacgcagcc gcaaagaacg gctatacacc actgcacatc
1861 gcggccaaga agaaccagat ggacatagcc acgtccctgc tggagtacgg tgctgatgca
1921 aacgcggtta cccggcaagg gattgcgtcc gtccatcttg cggcacagga agggcacgtg
1981 gacatggtgt cgctgctcct gagtagaaac gcgaatgtca acctgagcaa taagagcggt
2041 ctcaccccac tccacctggc tgctcaagaa gaccgagtga atgtggccga ggtccttgtc
2101 aaccaggggg cccatgtgga tgctcagaca aagatgggct acaccccgct ccatgtgggc
2161 tgtcactatg gaaatatcaa aatagtcaat tttctgctgc agcattctgc aaaagttaat
2221 gccaagacga agaatggata cacagcactg caccaggctg ctcagcaggg ccacacgcat
2281 atcatcaatg tcttgcttca gaacaacgcc tcccccaatg aactcactgt gaatgggaac
2341 acagctctgg ccatcgcccg gcgccttggt tacatctcgg tggttgacac actgaaggtc
2401 gtgacggagg aaattatgac caccactacc atcacggaga agcacaaaat gaatgtccca
2461 gaaacgatga atgaagtcct cgatatgtca gacgatgaag taaggaaagc cagcgccccc
2521 gaaaagctca gtgatgggga atatatctca gacggtgaag aaggtgataa atgcacatgg
2581 ttcaaaattc ccaaagtaca ggaggttttg gtgaaaagtg aagatgccat cacaggggac
2641 actgacaagt atctcgggcc acaggacctt aaggagctag gtgatgactc cctgccagca
2701 gaaggttacg taggcttcag tcttggagcc cgttctgcca gcctccgctc cttcagttcg
2761 gataggtcct acaccttgaa cagaagctcc tacgcaaggg acagcatgat gatagaggaa
2821 cttctggtac catccaaaga gcagcacctg acgttcacga gggagtttga ttctgactcc
2881 ctcagacact acagttgggc agcggacacg ttagataatg tgaacctggt ctcaagcccg
2941 gtgcattctg ggttttctggt tagctttatg gtggacgcga gaggggggctc catgcgagga
3001 agccgccacc acgggatgcg gatcatcatc cctccgcgaa agtgtacggc ccccacccgc
3061 atcacgtgcc gcctggtaaa gagacataaa ctggccaacc cacccccccat ggtggaagga
3121 gagggattag ccagtaggct ggtagaaatg ggtcctgcgg gggcacaatt tttaggcccc
3181 gtcattgtgg aaatccctca ttttgggtcc atgaggggga aggagagaga acttatcgtc
3241 cttcggagcg agaacggaga gacctggaag gaacatcagt ttgacagtaa aaacgaagac
3301 ctcgcggagc ttctcaatgg catggatgaa gaactcgaca gcccggaaga gttgggtaca
3361 aagcgcatct gcagaattat cacaaaggat ttcccccagt attttgccgt ggtttcccgg
3421 attaagcagg aaagcaacca gatcggtcct gagggtggga ttctgagcag caccaccgtg
3481 cccctcgtcc aggcctcctt cccagagggc gccttaacca agaggatccg tgtgggtctc
3541 caggctcagc ccgtgccaga ggaaacggta aaaaaaatcc ttgggaacaa agcaacattt
3601 agcccaattg tcacggtaga gccgaggaga aggaagttcc ataagccgat caccatgacc
3661 attccggtgc ccccgccctc gggagaaggc gtgtccaatg ggtacaaggg ggatgccacg

```
3721 cccaacctgc ggctcctctg cagcatcaca ggaggcacct caccagctca atgggaagac
3781 atcacaggaa caacccctct gacgttcata aaggattgtg tgtctttcac aaccaacgtt
3841 tcagccagat tctggctggc ggactgccat caggtgttag agaccgtagg gctagcctcc
3901 cagctgtaca gagagctgat atgcgttccc tacatggcca agttcgttgt gtttgccaaa
3961 acaaacgacc cggtggagtc ctcgctgagg tgcttctgta tgacagacga cagggtggac
4021 aaaaccctgg agcagcagga gaacttcgag gaggttgcca gaagcaaaga cattgaggtt
4081 ctggaaggaa agcccatcta cgttgattgc tatggaaacc tggcccctct gaccaaagga
4141 ggacagcagc ttgtttttaa cttttattct ttcaaagaaa acagactgcc attttccatc
4201 aagatcagag acaccagtca agagccctgt ggccgcctgt ctttcctgaa ggagccaaag
4261 acaacaaagg gattacccca aacagctgtt tgcaacttaa atattactct gccggcacat
4321 aaaaaggctg agaaggcaga cagacgccag agctttgcct ccctagcttt acgtaagcgc
4381 tacagctact tgactgaacc cagcatgagt ccgcagagtc cttgtgagcg gacggatatc
4441 aggatggcga tagtagccga tcacctggga cttagttgga cagagctggc aagggaactg
4501 aatttttcag tggatgaaat caaccaaata cgtgtggaaa atcccaattc tttaatttct
4561 cagagcttca tgttattaaa aaagtgggtg accagagacg gaaagaatgc cacaactgat
4621 gccttaactt cggtcttaac gaagattaac cggatagaca ttgtaactct gctggaagga
4681 ccaatatttg attatgggaa tatttcaggc accagaagct ttgcagatga aaacaatgtt
4741 ttccatgacc cagttgatgg ttggcagaac gagacgccaa gtggaagcct agagtcccca
4801 gcgcaagctc gaagactaac tggtgggtta ctggaccgtc tggatgacag ctctgaccag
4861 gctcgggatt ctattacctc atacctcacg ggagaacctg ggaagatcga agcaaatgga
4921 aaccacacag cggaagtcat tccagaagca aaggcaaaac cctacttccc ggaatcccaa
4981 aacgatatag ggaaacagag catcaaggag aacctgaaac caaaaacaca cggatgtggt
5041 cgcactgagg aaccagtgtc gcccctcaca gcctaccaga aatctctgga agaaaccagc
5101 aagcttgtca tagaagacgc acctaaaccc tgtgtgcctg tcggcatgaa aaagatgacc
5161 aggactacgg ctgacggcaa agccaggctc aacctccagg aagaagaggg gtccaccagg
5221 tcagagccta agcagggaga aggctataag gtgaagacga agaaggaaat ccggaacgtg
5281 gagaagaaaa cccactag
```

//

36. ANK3 (*Mus musculus*) Protein sequence

```
   1 mseepkekpa kpahrkrkgk ksdanasylr aaraghleka ldyiknqvdv nicnqnglna
  61 lhlaskeghv evvsellqre anvdaatkkg ntalhiasla gqaevvkvlv tnganvnaqs
 121 qngftplyma aqenhlevvr flldngasqs latedgftpl avalqqghdq vvslllendt
 181 kgkvrlpalh iaarkddtka aalllqndtn advesksgft plhiaahygn invatlllnr
 241 aaavdftarn ditplhvask rgnanmvkll ldrgakidak trdgltplhc garsgheqvv
 301 emlldrsapi lsktknglsp lhmatqgdhl ncvqlllqhn vpvddvtndy ltalhvaahc
 361 ghykvakvll dkkaspnaka lngftplhia ckknrirvme lllkhgasiq avtesgltpi
 421 hvaafmghvn ivsqlmhhga spnttnvrge talhmaarsg qaevvrylvq dgaqveakak
 481 ddqtplhisa rlgkadivqq llqqgaspna attsgytplh laareghedv aaflldhgas
 541 lsittkkgft plhvaakygk levaslllqk saspdaagks gltplhvaah ydnqkvalll
 601 ldqgasphaa akngytplhi aakknqmdia tslleygada navtrqgias vhlaaqeghv
 661 dmvslllsrn anvnlsnksg ltplhlaaqe drvnvaevlv nqgahvdaqt kmgytplhvg
 721 chygnikivn fllqhsakvn aktkngytal hqaaqqghth iinvllqnna spneltvngn
 781 talaiarrlg yisvvdtlkv vteeimtttt itekhkmnvp etmnevldms ddevrkasap
 841 eklsdgeyis dgeegdkctw fkipkvqevl vksedaitgd tdkylgpqdl kelgddslpa
```

901 egyvgfslga rsaslrsfss drsytlnrss yardsmmiee llvpskeqhl tftrefdsds
961 lrhyswaadt ldnvnlvssp vhsgflvsfm vdarggsmrg srhhgmriii pprkctaptr
1021 itcrlvkrhk lanpppmveg eglasrlvem gpagaqflgp viveiphfgs mrgkereliv
1081 lrsengetwk ehqfdskned laellngmde eldspeelgt kricriitkd fpqyfavvsr
1141 ikqesnqigp eggilssttv plvqasfpeg altkrirvgl qaqpvpeetv kkilgnkatf
1201 spivtveprr rkfhkpitmt ipvpppsgeg vsngykgdat pnlrllcsit ggtspaqwed
1261 itgttpltfi kdcvsfttnv sarfwladch qvletvglas qlyrelicvp ymakfvvfak
1321 tndpvesslr cfcmtddrvd ktleqqenfe evarskdiev legkpiyvdc ygnlapltkg
1381 gqqlvfnfys fkenrlpfsi kirdtsqepc grlsflkepk ttkglpqtav cnlnitlpah
1441 kkaekadrrq sfaslalrkr ysyltepsms pqspcertdi rmaivadhlg lswtelarel
1501 nfsvdeinqi rvenpnslis qsfmllkkwv trdgknattd altsvltkin ridivtlleg
1561 pifdygnisg trsfadennv fhdpvdghps fqveletpmg lywtppnpfq qddhfsdiss
1621 iespfrtpsr lsdglvpsqg niehptggpp vvtaedtsle dskmddsvtv tdpadpldvd
1681 esqlkdlcqs ecaqcwasvp gipndgrqae plrpqtrkvg msseqqekgk sgpdeevted
1741 kvkslfediq leeveaeemt edqqqamlnr vqraelamss lagwqnetps gslespaqar
1801 rltgglldrl ddssdqards itsyltgepg kieangnhta evipeakakp yfpesqndig
1861 kqsikenlkp kthgcgrtee pvspltayqk sleetsklvi edapkpcvpv gmkkmtrtta
1921 dgkarlnlqe eegstrsepk qgegykvktk keirnvekkt h

//

37. FGFR2 (*Homo sapiens*) cDNA sequence

1 atggtcagct ggggtcgttt catctgcctg gtcgtggtca ccatggcaac cttgtccctg
61 gcccggccct ccttcagttt agttgaggat accacattag agccagaaga gccaccaacc
121 aaataccaaa tctctcaacc agaagtgtac gtggctgcgc caggggagtc gctagaggtg
181 cgctgcctgt tgaaagatgc cgccgtgatc agttggacta aggatggggt gcacttgggg
241 cccaacaata ggacagtgct tattggggag tacttgcaga taaagggcgc cacgcctaga
301 gactccggcc tctatgcttg tactgccagt aggactgtag acagtgaaac ttggtacttc
361 atggtgaatg tcacagatgc catctcatcc ggagatgatg aggatgacac cgatggtgcg
421 gaagattttg tcagtgagaa cagtaacaac aagagagcac catactggac caacacagaa
481 aagatggaaa agcggctcca tgctgtgcct gcggccaaca ctgtcaagtt tcgctgccca
541 gccgggggga acccaatgcc aaccatgcgg tggctgaaaa acgggaagga gtttaagcag
601 gagcatcgca ttggaggcta caaggtacga aaccagcact ggagcctcat tatggaaagt
661 gtggtcccat ctgacaaggg aaattatacc tgtgtggtgg agaatgaata cgggtccatc
721 aatcacacgt accacctgga tgttgtggag cgatcgcctc accggcccat cctccaagcc
781 ggactgccgg caaatgcctc cacagtggtc ggaggagacg tagagtttgt ctgcaaggtt
841 tacagtgatg cccagcccca catccagtgg atcaagcacg tggaaaagaa cggcagtaaa
901 tacgggcccg acgggctgcc ctacctcaag gttctcaagg ccgccggtgt taacaccacg
961 gacaaagaga ttgaggttct ctatattcgg aatgtaactt ttgaggacgc tggggaatat
1021 acgtgcttgg cgggtaattc tattgggata tcctttcact ctgcatggtt gacagttctg
1081 ccagcgcctg gaagagaaaa ggagattaca gcttccccag actacctgga gatagccatt
1141 tactgcatag gggtcttctt aatcgcctgt atggtggtaa cagtcatcct gtgccgaatg
1201 aagaacacga ccaagaagcc agacttcagc agccagccgg ctgtgcacaa gctgaccaaa
1261 cgtatccccc tgcggagaca ggtaacagtt tcggctgagt ccagctcctc catgaactcc
1321 aacacccccgc tggtgaggat aacaacacgc ctctcttcaa cggcagacac ccccatgctg
1381 gcaggggtct ccgagtatga acttccagag gacccaaaat gggagtttcc aagagataag

1441 ctgacactgg gcaagcccct gggagaaggt tgctttgggc aagtggtcat ggcggaagca
1501 gtgggaattg acaaagacaa gcccaaggag gcggtcaccg tggccgtgaa gatgttgaaa
1561 gatgatgcca cagagaaaga cctttctgat ctggtgtcag agatggagat gatgaagatg
1621 attgggaaac acaagaatat cataaatctt cttggagcct gcacacagga tgggcctctc
1681 tatgtcatag ttgagtatgc ctctaaaggc aacctccgag aatacctccg agcccggagg
1741 ccacccggga tggagtactc ctatgacatt aaccgtgttc ctgaggagca gatgaccttc
1801 aaggacttgg tgtcatgcac ctaccagctg gccagaggca tggagtactt ggcttcccaa
1861 aaatgtattc atcgagattt agcagccaga aatgttttgg taacagaaaa caatgtgatg
1921 aaaatagcag actttggact cgccagagat atcaacaata tagactatta caaaaagacc
1981 accaatgggc ggcttccagt caagtggatg gctccagaag ccctgtttga tagagtatac
2041 actcatcaga gtgatgtctg gtccttcggg gtgttaatgt gggagatctt cactttaggg
2101 ggctcgccct acccagggat tcccgtggag gaacttttta agctgctgaa ggaaggacac
2161 agaatggata agccagccaa ctgcaccaac gaactgtaca tgatgatgag ggactgttgg
2221 catgcagtgc cctcccagag accaacgttc aagcagttgg tagaagactt ggatcgaatt
2281 ctcactctca caaccaatga ggaatacttg gacctcagcc aacctctcga acagtattca
2341 cctagttacc ctgacacaag aagttcttgt tcttcaggag atgattctgt ttttctccca
2401 gaccccatgc cttacgaacc atgccttcct cagtatccac acataaacgg cagtgttaaa
2461 acatga

//

38. FGFR2 (*Homo sapiens*) Protein sequence

1 mvswgrficl vvvtmatlsl arpsfslved ttlepeeppt kyqisqpevy vaapgeslev
61 rcllkdaavi swtkdgvhlg pnnrtvlige ylqikgatpr dsglyactas rtvdsetwyf
121 mvnvtdaiss gddeddtdga edfvsensnn krapywtnte kmekrlhavp aantvkfrcp
181 aggnpmptmr wlkngkefkq ehriggykvr nqhwslimes vvpsdkgnyt cvveneygsi
241 nhtyhldvve rsphrpilqa glpanastvv ggdvefvckv ysdaqphiqw ikhvekngsk
301 ygpdglpylk vlkaagvntt dkeievlyir nvtfedagey tclagnsigi sfhsawltvl
361 papgrekeit aspdyleiai ycigvfliac mvvtvilcrm knttkkpdfs sqpavhkltk
421 riplrrqvtv saessssmns ntplvrittr lsstadtpml agvseyelpe dpkwefprdk
481 ltlgkplgeg cfgqvvmaea vgidkdkpke avtvavkmlk ddatekdlsd lvsememmkm
541 igkhkniinl lgactqdgpl yviveyaskg nlreylrarr ppgmeysydi nrvpeeqmtf
601 kdlvsctyql argmeylasq kcihrdlaar nvlvtennvm kiadfglard innidyykkt
661 tngrlpvkwm apealfdrvy thqsdvwsfg vlmweiftlg gspypgipve elfkllkegh
721 rmdkpanctn elymmmrdcw havpsqrptf kqlvedldri ltlttneeyl dlsqpleqys
781 psypdtrssc ssgddsvfsp dpmpyepclp qyphingsvk t

//

39. FGFR2 (*Mus musculus*) cDNA sequence

1 atggtcagct gggggcgctt catctgcctg gtcttggtca ccatggcaac cttgtccctg
61 gcccggcccc ccttcagttt agttgaggat accactttag aaccagaagg agcaccgtac
121 tggaccaaca ccgagaagat ggagaagcgg ctccacgctg tccctgccgc caacactgtg
181 aagttccgct gtccggctgg ggggaatcca acgcccacaa tgaggtggtt aaaaaacggg
241 aaggagttta agcaggagca tcgcattgga ggctataagg tacgaaacca gcactggagc
301 cttattatgg aaagtgtggt cccgtcagac aaaggcaact acacctgcct ggtggagaat
361 gaatacgggt ccatcaacca cacctaccac ctcgatgtcg ttgaacggtc accacaccgg
421 cccatcctcc aagctggact gcctgcaaat gcctccacgg tggtcggagg ggatgtggag

```
 481 tttgtctgca aggtttacag cgatgcccag ccccacatcc agtggatcaa gcacgtggaa
 541 aagaacggca gtaaatacgg gcctgatggg ctgccctacc tcaaggtcct gaagcactcg
 601 gggataaata gctccaatgc agaagtgctg gctctgttca atgtgacgga gatggatgct
 661 ggggaatata tatgtaaggt ctccaattat atagggcagg ccaaccagtc tgcctggctc
 721 actgtcctgc ccaaacagca agcgcctgtg agagagaagg agatcacggc ttccccagat
 781 tatctggaga tagctattta ctgcataggg gtcttcttaa tcgcctgcat ggtggtgaca
 841 gtcatctttt gccgaatgaa gaccacgacc aagaagccag acttcagcag ccagccagct
 901 gtgcacaagc tgaccaagcg catccccctg cggagacagg taacagtttc ggccgagtcc
 961 agctcctcca tgaactccaa caccccgctg gtgaggataa caacgcgtct gtcctcaaca
1021 gcggacaccc cgatgctagc aggggtctcc gagtatgagt tgccagagga tccaaagtgg
1081 gaattcccca gagataagct gacgctgggc aaaccccctgg gggaaggttg cttcgggcaa
1141 gtagtcatgg ctgaagcagt gggaatcgat aaagacaaac ccaaggaggc ggtcaccgtg
1201 gcagtgaaga tgttgaaaga tgatgccaca gagaaggacc tgtctgatct ggtatcagag
1261 atggagatga tgaagatgat tgggaaacat aagaacatta tcaacctcct gggggcctgc
1321 acgcaggatg gacctctcta cgtcatagtt gaatatgcat cgaaaggcaa cctccgggaa
1381 tacctccgag cccggaggcc acctggcatg gagtactcct atgacattaa ccgtgtcccc
1441 gaggagcaga tgaccttcaa ggacttggtg tcctgcacct accagctggc tagaggcatg
1501 gagtacttgg cttcccaaaa atgtatccat cgagatttgg ctgccagaaa cgtgttggta
1561 acagaaaaca atgtgatgaa gatagcagac tttggcctgg ccagggatat caacaacata
1621 gactactata aaaagaccac aaatgggcga cttccagtca agtggatggc tcctgaagcc
1681 ctttttgata gagtttacac tcatcagagc gatgtctggt ccttcggggt gttaatgtgg
1741 gagatcttta ctttaggggg ctcaccctac ccagggattc ccgtggagga acttttaag
1801 ctgctcaaag agggacacag gatggacaag cccaccaact gcaccaatga actgtacatg
1861 atgatgaggg attgctggca tgctgtaccc tcacagagac ccacattcaa gcagttggtc
1921 gaagacttgg atcgaattct gactctcaca accaatgagg aatacttgga tctcacccag
1981 cctctcgaac agtattctcc tagttacccc gacacaagga gctcttgttc ttcaggggac
2041 gattctgtgt tttctccaga ccccatgcct tatgaaccct gtctgcctca gtatccacac
2101 ataaacggca gtgttaaaac atga
```

//

40. FGFR2 (Mus musculus) Protein sequence

```
   1 mvswgrficl vlvtmatlsl arpsfslved ttlepegapy wtntekmekr lhavpaantv
  61 kfrcpaggnp tptmrwlkng kefkqehrig gykvrnqhws limesvvpsd kgnytclven
 121 eygsinhtyh ldvversphr pilqaglpan astvvggdve fvckvysdaq phiqwikhve
 181 kngskygpdg lpylkvlkhs ginssnaevl alfnvtemda geyickvsny igqanqsawl
 241 tvlpkqqapv rekeitaspd yleiaiycig vfliacmvvt vifcrmkttt kkpdfssqpa
 301 vhkltkripl rrqvtvsaes sssmnsntpl vrittrlsst adtpmlagvs eyelpedpkw
 361 efprdkltlg kplgegcfgq vvmaeavgid kdkpkeavtv avkmlkddat ekdlsdlvse
 421 memmkmigkh kniinllgac tqdgplyviv eyaskgnlre ylrarrppgm eysydinrvp
 481 eeqmtfkdlv sctyqlargm eylasqkcih rdlaarnvlv tennvmkiad fglardinni
 541 dyykkttngr lpvkwmapea lfdrvythqs dvwsfgvlmw eiftlggspy pgipveelfk
 601 llkeghrmdk ptnctnelym mmrdcwhavp sqrptfkqlv edldriltlt tneeyldltq
 661 pleqyspsyp dtrsscssgd dsvfspdpmp yepclpqyph ingsvkt
```

//

41. CES1 (*Homo sapiens*) cDNA sequence

```
   1 atgtggctcc gtgcctttat cctggccact ctctctgctt ccgcggcttg ggcagggcat
  61 ccgtcctcgc cacctgtggt ggacaccgtg catggcaaag tgctggggaa gttcgtcagc
 121 ttagaaggat ttgcacagcc tgtggccatt ttcctgggaa tccctttgc caagccgcct
 181 cttggacccc tgaggtttac tccaccgcag cctgcagaac catggagctt tgtgaagaat
 241 gccacctcgt accctcctat gtgcacccaa gatcccaagg cggggcagtt actctcagag
 301 ctatttacaa accgaaagga gaacattcct ctcaagcttt ctgaagactg tctttacctc
 361 aatatttaca ctcctgctga cttgaccaag aaaaacaggc tgccggtgat ggtgtggatc
 421 cacggagggg ggctgatggt gggtgcggca tcaacctatg atgggctggc ccttgctgcc
 481 catgaaaacg tggtggtggt gaccattcaa tatcgcctgg gcatctgggg attcttcagc
 541 acaggggatg aacacagccg ggggaactgg ggtcacctgg accaggtggc tgccctgcgc
 601 tgggtccagg acaacattgc cagctttgga gggaacccag gctctgtgac catctttgga
 661 gagtcagcgg gaggagaaag tgtctctgtt cttgttttgt ctccattggc caagaacctc
 721 ttccaccggg ccatttctga gagtggcgtg gccctcactt ctgttctggt gaagaaaggt
 781 gatgtcaagc ccttggctga gcaaattgct atcactgctg ggtgcaaaac caccacctct
 841 gctgtcatgg ttcactgcct gcgacagaag acggaagagg agctcttgga gacgacattg
 901 aaaatgaaat tcttatctct ggacttacag ggagacccca gagagagtca acccctctg
 961 ggcactgtga ttgatgggat gctgctgctg aaaacacctg aagagcttca agctgaaagg
1021 aatttccaca ctgtccccta catggtcgga attaacaagc aggagtttgg ctggttgatt
1081 ccaatgcagt tgatgagcta tccactctcc gaagggcaac tggaccagaa gacagccatg
1141 tcactcctgt ggaagtccta tccccttgtt tgcattgcta aggaactgat tccagaagcc
1201 actgagaaat acttaggagg aacagacgac actgtcaaaa agaaagacct gttcctggac
1261 ttgatagcag atgtgatgtt tggtgtccca tctgtgattg tggcccggaa ccacagagat
1321 gctggagcac ccacctacat gtatgagttt cagtaccgtc caagcttctc atcagacatg
1381 aaacccaaga cggtgatagg agaccacggg gatgagctct tctccgtctt tggggcccca
1441 tttttaaaag agggtgcctc agaagaggag atcagactta gcaagatggt gatgaaattc
1501 tgggccaact ttgctcgcaa tggaaacccc aatggggaag ggctgcccca ctggccagag
1561 tacaaccaga aggaagggta tctgcagatt ggtgccaaca cccaggcggc ccagaagctg
1621 aaggacaaag aagtagcttt ctggaccaac ctctttgcca agaaggcagt ggagaagcca
1681 ccccagacag aacacataga gctgtga
```

//

42. CES1 (*Homo sapiens*) Protein sequence

```
   1 mwlrafilat lsasaawagh pssppvvdtv hgkvlgkfvs legfaqpvai flgipfakpp
  61 lgplrftppq paepwsfvkn atsyppmctq dpkagqllse lftnrkenip lklsedcly1
 121 niytpadltk knrlpvmvwi hggglmvgaa stydglalaa henvvvvtiq yrlgiwgffs
 181 tgdehsrgnw ghldqvaalr wvqdniasfg gnpgsvtifg esaggesvsv lvlsplaknl
 241 fhraisesgv altsvlvkkg dvkplaeqia itagckttts avmvhclrqk teeellettl
 301 kmkflsldlq gdpresqpll gtvidgmlll ktpeelqaer nfhtvpymvg inkqefgwli
 361 pmqlmsypls egqldqktam sllwksyplv ciakelipea tekylggtdd tvkkkdlfld
 421 liadvmfgvp svivarnhrd agaptymyef qyrpsfssdm kpktvigdhg delfsvfgap
 481 flkegaseee irlskmvmkf wanfarngnp ngeglphwpe ynqkegylqi gantqaaqkl
 541 kdkevafwtn lfakkavekp pqtehiel
```

//

43. CES1 (*Mus musculus*) cDNA sequence

```
   1 atgtggctct gtgctttgag tctgatctct ctcactgctt gcttgagtct gggacaccca
```

```
     61 tccttaccgc ctgtggtaca caccgttcat ggcaaagtcc tggggaagta tgtcaccttа
    121 gaaggattct cacagcctgt ggccgtcttc ctgggagtcc cctttgccaa gccccctctt
    181 ggatctctga ggtttgctcc accagagcct gcagagccct ggagcttcgt gaagcacacc
    241 acttcctacc ctcctttgtg ctaccaaaac ccagaggcag cattgaggct cgctgagcgc
    301 ttcaccaacc aaaggaagat cattccccac aaattttctg aggactgtct ctacctcaac
    361 atttatactc ctgctgactt aacacagaac agcaggttgc ccgtgatggt gtggatacat
    421 ggaggtggac ttgtgataga tggagcatca acctatgatg gagtgccccc ggctgtccat
    481 gaaaatgtgg ttgtagtggt cattcagtat cgcctgggca tctggggatt cttcagcaca
    541 gaggatgaac acagccgggg gaactggggt cacttggacc aggtggctgc actacattgg
    601 gtccaagaca acattgccaa ctttgggggc aacccaggat ctgtgactat cttcggcgag
    661 tcagcaggag gtgaaagtgt ctctgttctt gtgttaagcc cactggccaa gaacctcttc
    721 cacagggcca tcgctcagag tagtgtcatt ttcaatcctt gccttttttgg gagagctgcc
    781 agacccttgg ctaagaaaat tgctgctctt gctggctgta aaaccaccac ctccgctgcc
    841 atggttcact gcctgcgcca gaagactgaa gatgagctct tggaggtctc actgaaaatg
    901 aaatttggga ctgttgattt tcttggagac cccagagaga gctatccctt cctccctact
    961 gtgattgatg gagtgttgct gccaaaggca ccagaagaga ttctggctga gaagagtttc
   1021 aacactgtcc cctacatggt gggcatcaac aagcatgagt ttggctggat cattccaatg
   1081 tttttggact tcccactctc tgaaagaaaa ctggaacaga agacagctgc atccatcctg
   1141 tggcaggcct acccaattct taacatctct gaaaagctga ttccagcagc tattgaaaag
   1201 tatttaggag ggacagaaga ccctgccaca atgacagacc tgttcctgga cttgattgga
   1261 gacattatgt tcggtgtccc atctgtaatc gtgtcccgta gtcacagaga tgctggagcc
   1321 ccaacctaca tgtatgaata tcagtatcgc ccaagttttg tatcagacga tagaccccag
   1381 gaattgttag gagaccacgc tgatgaactc ttttctgtat ggggagcccc gtttttaaaa
   1441 gagggtgctt cagaagaaga gatcaacctc agcaacatgg tgatgaaatt ctgggccaac
   1501 tttgctcgga atgggaaccc taatggtgaa gggctgcctc attggccaga atatgaccag
   1561 aaggaaggat accttcagat tggagtccca gcacaggcag cccataggct gaaagacaag
   1621 gaagtggact tttggactga gctcagagcc aaggaaacag cagagaggtc atcccatagg
   1681 gaacatgttg aactgtga
//
```

44. CES1 (Mus musculus) Protein sequence

```
      1 mwlcalslis ltaclslghp slppvvhtvh gkvlgkyvtl egfsqpvavf lgvpfakppl
     61 gslrfappep aepwsfvkht tsypplcyqn peaalrlaer ftnqrkiiph kfsedclyln
    121 iytpadltqn srlpvmvwih ggglvidgas tydgvplavh envvvvviqy rlgiwgffst
    181 edehsrgnwg hldqvaalhw vqdnianfgg npgsvtifge saggesvsvl vlsplaknlf
    241 hraiaqssvi fnpclfgraa rplakkiaal agcktttsaa mvhclrqkte dellevslkm
    301 kfgtvdflgd presypflpt vidgvllpka peeilaeksf ntvpymvgin khefgwiipm
    361 fldfplserk leqktaasil wqaypilnis eklipaaiek ylggtedpat mtdlfldlig
    421 dimfgvpsvi vsrshrdaga ptymyeyqyr psfvsddrpq ellgdhadel fsvwgapflk
    481 egaseeeinl snmvmkfwan farngnpnge glphwpeydq kegylqigvp aqaahrlkdk
    541 evdfwtelra ketaersshr ehvel
//
```